

56

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
14 March 2002 (14.03.2002)

PCT

(10) International Publication Number  
**WO 02/20651 A2**

(51) International Patent Classification<sup>7</sup>: **C08K**

(21) International Application Number: PCT/US01/27638

(22) International Filing Date:  
6 September 2001 (06.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/230,258 6 September 2000 (06.09.2000) US  
60/257,130 20 December 2000 (20.12.2000) US

(71) Applicant: ZYMOGENETICS, INC. [US/US]; 1201  
Eastlake Avenue East, Seattle, WA 98102 (US).

(72) Inventors: HOLLOWAY, James, L.; 835 NE 89th Street,  
Seattle, WA 98115 (US). LOK, Si; 806 NW 52nd Street,  
Seattle, WA 98107 (US).

(74) Agent: JONES, Phillip, B., C.; ZymoGenetics, Inc., 1201  
Eastlake Avenue East, Seattle, WA 98102 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



**WO 02/20651 A2**

(54) Title: HUMAN PHERMONE POLYPEPTIDE

(57) Abstract: Human pheromones may be used to alleviate anxiety, promote beneficial moods, and to alter hypothalamic functions, such as satiety, energy balance, and reproductive biology. The present invention provides a new human pheromone polypeptide, designated "ZHUMP-2."

## HUMAN PHERMONE POLYPEPTIDE

5

### TECHNICAL FIELD

The present invention relates generally to a new gene that encodes a new human protein. In particular, the present invention relates to a novel protein, designated "ZHMUP-2," and to nucleic acid molecules encoding ZHMUP-2.

### BACKGROUND OF THE INVENTION

Olfaction is an ancient sense, rudiments of which can be found in the most primitive single-celled organisms (see generally, Tirindelli *et al.*, *TINS* 11:482 (1998); Keverne, *Science* 286:716 (1999); Liman, *Current Opinion in Neurobiology* 6:487 (1996); Buck, *Cell* 65:175 (2000)). The importance of this sense is exemplified by the fact that humans are capable of perceiving thousands of discrete odors, and that more than 1% of the genes in the human genome are devoted to olfaction. Olfaction has an aesthetic component that is capable of invoking emotion and memory leading to measured thoughts and response to the everyday environment. However, in some species, a diverse class of molecules, generally referred to as pheromones, can elicit innate and stereotyped behaviors that are likely to result from non-conscious perception.

The term "pheromone" was introduced into the scientific literature in 1959 by Karlson and Luscher, *Nature* 183:55 (1959), who defined pheromone as "a substance secreted by an animal to the outside of that individual, which then elicits some behavioral or developmental response in the latter." At present, the majority of the identified pheromones are from insects. Many insect species produce potent volatile chemical compounds, which attract potential mates over long distances (Kaissling, *Ann. Rev. Neurosci.* 9:121 (1986); Masson and Mustaparta, *Physiol. Rev.* 70:199 (1990)). Synthetic versions of certain pheromones are used as chemo-attractants to control insect pests.

Members of the animal kingdom are also known to produce pheromones for intra-species communication. F-prostaglandins and steroids, for example, have been shown to induce sperm production and mating in fish (Stacey and Sorensen, *Can. J. Zool.* 64:2412 (1986); Sorensen *et al.*, *Biol. Reprod.* 39:1039 (1988)). In reptiles, a

family of dieneackeronone-related steroidal esters was characterized from the crocodilian paracloacal gland secretions, which are thought to contain pheromones that may play a role in nesting and mating activities (Whyte *et al.*, *Proc. Nat'l Acad. Sci. (USA)* 96:12246 (1999); Yang *et al.*, *Proc. Nat'l Acad. Sci. (USA)* 96:12251 (1999)). A series of nonvolatile saturated and monosaturated long-chain methyl ketones, and compounds containing squalene were shown to induce courtship behavior in garter snakes (Mason *et al.*, *Science* 293:290 (1989)). Recently, a proteinaceous pheromone affecting female receptivity was isolated from a terrestrial salamander, and a peptide with female-attracting activity was identified in newts (Kikuyama *et al.*, *Science* 267:1643 (1995); Rollmann *et al.*, *Science* 285:1907 (1999)).

Mammalian pheromones have also been described. In mammals, the two pathways of olfactory perception are mediated by anatomically and functionally distinct sensory organs. The main olfactory epithelium recognizes everyday odorants, whereas the vomeronasal organ perceives pheromones (see, for example, Buck, *Cell* 65:175 (2000); Liman, *Current Opinion in Neurobiology* 6:487 (1996); Tirindelli *et al.*, *TINS* 11:482 (1998); Keverne, *Science* 286:716 (1999)). The main olfactory epithelium and the neuroepithelium of the vomeronasal organ contain sensory neurons that project axons to the brain (Belluscio *et al.*, *Cell* 97:209 (1999); Rodriguez *et al.*, *Cell* 97:199 (1999)). Sensory inputs from the main olfactory epithelium ultimately reach multiple regions of the brain, including the frontal cortex, which is believed to process the conscious perception of odors. In contrast, pheromone derived signals from the vomeronasal organ bypass higher cognitive centers and are processed directly in regions of the amygdala and hypothalamus that have been implicated in the regulation of innate behavior, reproductive physiology, energy balance and other neuroendocrine responses.

Rodents provide useful experimental animals for studying pheromone action. The main vehicle of olfactory chemo-signals in the mouse is urine, which mediates a variety of behavioral and physiological responses. The role of saliva in sexual communication has also been demonstrated in mice (Marchlewska *et al.*, *J. Chem. Ecol.* 16:2817 (1990)). The endocrine effects primed by male mouse urine include: acceleration of female puberty onset, pregnancy block, attraction to females, aggression, estrus acceleration, and estrus synchronization. Pheromone signaling in mice is characterized by at least three components: (1) a special chemosensory organ, the vomeronasal organ; (2) volatile pheromone odorants; and (3) a high concentration of pheromone binding proteins in the male mouse urine. Volatile pheromone molecules in urine are bound to a group of carrier proteins known as the major urinary proteins (MUP). These proteins are believed to promote stability of the bound pheromone and to effect their sustained release from urine (Hurst *et al.*, *Anim. Behav.* 55:1289 (1988)).

A number of rodent volatile pheromones in mouse urine that bind with MUPs were recently characterized. Two major volatile constituents of the male rodent preputial gland, E,E-alpha-farnesene and E-beta-farnesene, were shown to attract females and to induce estrus (Jemiolo *et al.*, *Physiology & Behavior* 50:1119 (1991); Ma *et al.*,  
5 *Chem. Senses* 24:289 (1999)). Another urine component, 6-hydroxy-6-methyl-3-heptanone, is a pheromone that accelerates puberty in female mice (Novotny *et al.*, *Chemistry & Biology* 6:377 (1999)). Other volatile chemical compounds found in rodent urine, thiazole (2-sec-butyl-4,5-dihydrothiazole) and brevicomin (2,3-dehydro-exo-brevicomin), also function as attractants for females, inducers of estrous, and instigators  
10 of inter-male aggression (Jemiolo *et al.*, *Anim. Behav.* 33:1114 (1985); Novotny *et al.*, *Proc. Nat'l Acad. Sci. (USA)* 82:2059 (1985); Jemiolo *et al.*, *Proc. Nat'l Acad. Sci. (USA)* 83:4576 (1986); Hurst *et al.*, *Anim. Behav.* 55:1289 (1988); Novotny *et al.*, *Proc. R. Soc. Lond. B. Biol. Sci.* 266:2017 (2000)).

The MUPs, rodent pheromone carrier proteins, are members of the  
15 lipocalin family of extracellular proteins (see, for example, Flower, *FEBS Lett.* 354:7 (1994); Flower, *Biochem. J.* 318:1 (1996)). Together with the fatty-acid-binding proteins and the avidins, the lipocalins form part of a structural superfamily known as the calycins. Lipocalins are characterized by a single eight-stranded hydrogen-bonded anti-parallel  $\beta$ -barrel, which in some members encloses an internal ligand-binding-site  
20 (Lucke *et al.*, *Eur. J. Biochem.* 266:1210 (1999)). It is believed that one important function of the lipocalins is to control and to modulate the transport of small hydrophobic regulatory molecules between cells (Flower, *FEBS Lett.* 354:7 (1994)). Other portions of the protein are known to interact with cell-surface receptors or soluble macromolecules, which further add to the complex biological functions of these proteins.

25 Phylogenetic analysis of the lipocalins separates the family members into 13 monophyletic clades or groups (Ganformina *et al.*, *Mol. Biol. Evol.* 17:114 (2000)). The rodent MUPs comprise clade XIII. The odorant binding proteins, another class of lipocalins involved in olfaction perception, are in clade X (Ganformina *et al.*, *Mol. Biol. Evol.* 17:114 (2000)). A major difference in the three-dimensional structures of the  
30 MUPs and the odorant binding proteins lies in the ligand binding pocket (Böcskei *et al.*, *Nature* 360:186 (1992); Bianchet *et al.*, *Nat. Struct. Biol.* 3:934 (1996); Tegoni *et al.*, *Nat. Struct. Biol.* 3:863 (1996)). In both cases, the ligand binding pockets are lined by hydrophobic amino acid residues. However, the ligand binding pocket of the odorant binding protein contains a relatively large number of aromatic amino acids, which  
35 contribute to a smooth surface. In contrast, the corresponding region of the MUPs are rich in branch-chain amino acids, such as valine and leucine, providing a more complex binding surface that results in greater ligand specificity.



The murine MUPs are the products of a multi-gene family of approximately 35 genes and pseudogenes located on mouse chromosome 4 (Bishop *et al.*, *EMBI J* 1:615 (1982); Al-Shawi *et al.*, *J. Mol. Evol.* 29:302 (1989); Shi *et al.*, *Proc. Nat'l Acad. Sci. (USA)* 86:4584 (1989)). Murine MUPs are synthesized in at least six tissues (Shahan *et al.*, *Mol. Cell. Biol.* 7:1947 (1987)). MUP-I, -II, and -III are the most abundant MUPs expressed in the liver. MUP-II is also expressed in mammary gland. MUP-IV is expressed in the lachrymal and the parotid glands. MUP-V is expressed in the submaxillary, sublingual, and the lachrymal glands. MUP-VI is expressed in the parotids in BALB/c mice. Circulating MUP polypeptides are efficiently filtered by the kidney and are released into the urine along with their bound pheromone ligand at high concentrations that can approach 1 mM (1-5 mg/ml).

An important recent finding is that, in addition to being proteinaceous carriers of small volatile pheromones, murine MUPs without bound ligands have pheromone activity as shown by their ability to induce the acceleration of puberty in female mice (Mucignat-Caretta *et al.*, *J. Physiol.* 486:517 (1995)). Furthermore, a hexapeptide derived from the amino-terminus of murine MUP also is active in the assay (Clark *et al.*, *EMBO J.* 4:3159 (1985); Mucignat-Caretta *et al.*, *J. Physiol.* 486:517 (1995)). It is also reported that recombinant aphrodisin, a lipocalin family member found in vaginal discharge that can induce investigatory and copulatory responses in male hamsters, is active as a hamster pheromone in the apparent absence of a ligand (Macrides, *et al.*, *Physiol. Behav.* 33:633 (1984); Singer *et al.*, *J. Biol. Chem.* 261:13312 (1986); Henzel *et al.*, *J. Biol. Chem.* 263:16682 (1988); Singer and Macrides, *Chem. Senses* 15:199 (1990)). Pheromone activity of recombinant aphrodisin, however, is enhanced in the presence of organic extracts of hamster vaginal discharge suggesting that an as yet unidentified lipophilic ligand and the aphrodisin protein both contribute to the overall pheromone response (Singer and Macrides, *Chem. Senses* 15:199 (1990)). Polypeptides with pheromone activity are not without precedence. There are at least three reports of proteinaceous pheromones in amphibian species (Kikuyama *et al.*, *Science* 267:1643 (1995); Lebioda *et al.*, *Nature* 401:444 (1999); Rollmann *et al.*, *Science* 285:1907 (1999)).

It appears that the pheromone system has evolved to recognize and to respond to both the ligand and its carrier protein in hamster aphrodisin and some members of the murine MUP family. Consistent with this hypothesis is that many lipocalin proteins including the MUPs have regions on their surface, which are believed to interact with cell surface receptors and other regulatory molecules (Bocskei *et al.*, *Nature* 360:186 (1992)). Results from signal transduction experiments support the hypothesis that MUPs have a direct and independent role in pheromone recognition.

Two MUP ligands, brevicomin and dihydrothiazole, appear to activate only a small subset of neurons of the accessory olfactory bulb when compared with the ligand and the MUP carrier (Brennan *et al.*, *Neuro-Science* 90:1463 (1999)). Other evidence comes from rat  $\alpha$ -2-globulin, an orthologous protein to murine MUP. Recombinant  $\alpha$ -2-globulin was found to activate G-protein subtype  $G_o$ , whereas stimulation with the  $\alpha$ -2-globulin ligand alone resulted in the activation of G-protein,  $G_i$ , in vomeronasal organ membrane preparations (Krieger *et al.*, *J. Biol. Chem.* 274:4655 (1999)). Together, these results not only show that the MUPs and their ligands have independent pheromone activity, but that they can also activate distinct signaling pathways within the vomeronasal organ. The MUPs are also highly polymorphic proteins, and there is considerable genetic heterogeneity among the MUPs of different mouse strains (Robertson *et al.*, *Biochem. J.* 316:265 (1996); Robertson *et al.*, *Rapid Commun. Mass Spectrom.* 11:786 (1997)). To a nocturnal, burrowing animal, such as the mouse and rat, a pheromone recognition system that is in part mediated by a genetically encoded protein would allow for kin and individual recognition, and territorial demarcation.

Mammalian pheromone recognition is mediated by the vomeronasal organ, which resides within a blinded pouch in the septum of the nose (see, for example, Stensaas *et al.*, *J. Steroid. Biochem. Mol. Biol.* 39:553 (1991); Monti-Bloch *et al.*, *Annals New York Academy of Sciences* 30:373 (1998); Trindelli *et al.*, *TINS* 21:482 (1998); Keverne, *Science* 286:716 (1999)). Two distinct families of pheromone receptor genes, V1 and V2, are expressed in rodent vomeronasal neurons (Dulac and Axel, *Cell* 83:195 (1995); Herrada and Dulac, *Cell* 90:763 (1997); Matsunami and Buck, *Cell* 90:775 (1997); Ryba and Trindelli, *Neuron* 19:371 (1997); Dulac and Axel, *Chem. Senses* 23:467 (1998)). The V1 and V2 receptor genes comprised two novel families of seven-transmembrane domain G-protein coupled receptor proteins that are distinct from the odorant receptors expressed in the main olfactory epithelium or to other families of seven-transmembrane domain receptors (Buck and Axel, *Cell* 65:175 (1991)).

The V2 receptors are related to the metabotropic glutamate receptors, and have a large N-terminal domain that binds the ligand (O'Hara *et al.*, *Neuron* 11:41 (1993)). The V1 receptor ligand-binding pocket is formed from the transmembrane segments or by the peptide loops between the transmembrane segments. The different structure of the V1 and V2 receptor ligand binding pockets suggests these receptors recognize different types of ligands. Recent work of Krieger *et al.*, *J. Biol. Chem.* 274:4655 (1999), has provided experimental evidence in support of V1 receptors being activated by lipophilic volatile odorants, and V2 receptors interacting with proteinaceous pheromone components such as the MUPs. In this way, the dual recognition of a MUP and its ligand may be mediated separately by two distinct classes

of pheromone receptors. Thus, the pheromone response is apparently due to the collective signals from these two receptors.

Pheromone activities affecting sexual behavior or development have been reported in primates. Short-chain fatty acids found in vaginal secretions of rhesus monkeys can act as sex-attractants (Keverne and Michael, *J. Endocrinol.* 51:313 (1971)). Estradiol was reported as a pheromonal attractant of male rhesus monkeys (Michael *et al.*, *Nature* 218:746 (1968)). The removal of the vomeronasal organ in lower primates was shown to impair male sexual behavior consistent with the existence of a pheromone whose action is mediated through the vomeronasal organ (Aujard, *Physiol. Behav.* 62:1003 (1997)). From these findings, it would seem that primate sexual behavior is at least in part influenced by pheromones.

The existence of human pheromones, however, is controversial. Human reproductive behavior is largely independent of oestrous-promoting hormones. Maternal behavior may occur without pregnancy and sexual human behavior is also tempered by culture, learning and personal experience. Moreover, evolutionary enlargement of the human neocortex has enabled the rapid assimilation and integration of information from a number of senses. Hence, it has been argued that it is implausible that humans would be under significant behavior and endocrine regulation by pheromones. Nevertheless, the existence of human pheromones was first suggested by the observation that women living together can develop synchronized menstrual cycles under specific conditions (McClintock, *Nature* 291:244 (1971)). The causal agents were later attributed to odorless pheromone-like substances produced in female underarms (Stern and McClintock, *Nature* 392:177 (1998)). There are also reports suggesting short-chain fatty acids found in vaginal secretions isolated from vaginal secretion of sexually active human females can act as sex-attractants (Michael *et al.*, *Psychoneuroendocrinology* 1:153 (1975)); Sokolov *et al.*, *Archives of Sexual Behavior* 5:269 (1976)).

Much human pheromone research has centered on the 16-androstenes, which comprise a family of related steroids that have pheromone activity in animals. Androsterone (5- $\alpha$ -16-androst-16-en-3-one) and its alcohol form, androstenol (5- $\alpha$ -16-androst-16-en-3-ol) are porcine pheromones synthesized in the boar testes and submaxillary glands and, which induce recipient sows to adopt the mating stance (Reed and Melrose, *Br. Vet. J.* 130:61 (1974); Perry *et al.*, *Animal Production* 31:191 (1980)). These and other related 16-androstenes are also synthesized in human testes and believed by many investigators to have pheromone activity in humans (see, for example, Gower and Ruparelia, *J. Endocrinol.* 137:167 (1993); U.S. Patent No. 5,278,241; U.S. Patent No. 5,272,134; U.S. Patent No. 5,969,168; U.S. Patent No. 5,965,552). 5- $\alpha$ -16-androst-16-en-3-ol is the most abundant of the 16-androstenes in human urine.

Androsta-4,16-dien-3-one is the most abundant 16-androstene present in human semen, in male axillary hair and male axillary skin surfaces (Nixon *et al.*, *J. Steroid Biochem. Mol. Biol.* 29:505 (1988); Rennie *et al.*, In: *Chemical Signals in Vertebrates*, pages 55-60 (Oxford University Press 1990); Kwan *et al.*, *J. Steroid Biochem. Mol. Biol.* 43:549  
5 (1992)). Androstenes are also found in the human axillary sweat secreted by the apocrine glands, which are sites for pheromone production in lower animals (Brooksbank *et al.*, *Experientia* 30:864 (1994)).

Androsta-4,16-dien-3-one was reported to stimulate the human vomeronasal organ (Jennings-White, *Perfum. Flav.* 20:1 (1995); Monti-Bloch *et al.*,  
10 *Chem. Sens.* 23:114 (1998)). The same group later reported that the administration of androstadienone at picogram levels directly to the human female vomeronasal organ can significantly reduce discomfort and tension (Grosser *et al.*, *Psychoneuroendocrinology* 25:289 (2000)). While other studies also suggested that 16-androstenes and other putative pheromones may indeed alter human social behavior, there are also reports of  
15 negative and contradictory results (Filsinger *et al.*, *J. Comp. Psychol.* 98:219 (1984); Gustavson *et al.*, *Psychol.* 101:210 (1987); Cowley and Brooksbank, *J. Steroid Biochem. Molec. Biol.* 39:647 (1991); Gower and Ruparelia, *J. Endocrinol.* 137:167 (1993); Pause *et al.*, *Physiology & Behavior* 68:129 (1999)). The inconsistent findings may be due to different forms or formulation of the 16-androstenes used, or due to the subjectivity and  
20 the difficulties associated with human behavioral studies.

An alternative explanation is that a more robust reproducible human pheromone response to the androstrenes or other potential small chemical pheromones such as the estrenes (U.S. Patent Nos. 5,272,134, 5,278,141, and 5,994,568) may require the presence of a human lipocalin carrier protein. Such a protein carrier may by itself  
25 possess phermone activity, as in the case for the murine MUPs and hamster aphrodisin. Alternatively, the carrier protein may function indirectly by augmenting, stabilizing, or effecting the sustained release of phermone ligand. In addition to hamster aphrodisin and the rodent MUP, there have only been a few examples of characterized proteins associated with pheromones in mammals and none so far in humans. Booth and White  
30 reported a partially characterized porcine extracellular protein, pheromaxein, that binds androstenol and related steroids in boar submaxillary gland saliva (Booth and White, *J. Endrocr.* 118:47 (1988)). A salivary gland lipocalin, which binds 16-androstrenes, was later isolated from boar submaxillary gland (Marchese *et al.*, *Eur. J. Biochem.* 252:563 (1998)). A cDNA encoding this protein, termed sex-specific salivary lipocalin (SAL),  
35 was recently reported, and shown to encode a polypeptide with high homology to the murine MUPs (Loebel *et al.*, *Biochem. J.* 350:369 (2000)). It is not known whether boar SAL or pheromaxein has phermone activity itself, or only contributes to the phermone

activity of cognate androstene ligands. Human homologs of boar pheromaxin or SAL have not been characterized. However, a human lipocalin, apolipoprotein D, was recently found expressed in apocrine glands, and was shown to bind the axillary odorant, E-3-methyl-2-hexenoic acid (E-3M2H) (Zeng *et al.*, *Proc. Nat'l Acad. Sci. (USA)* 93:6626 (1996)). E-3M2H and its isomers are major odorants in the human axillary region. Studies have implicated axillary odors and secretions from both male and female in alterations of menstrual cycle (McClintock, *Nature* 291:244 (1971); Stern and McClintock, *Nature* 392:177 (1998)). Moreover, mood changes have been reported by human volunteers exposed to donor underarm odorants (Chen and Haviland-Jones, *Physiology & Behavior* 68:241 (1999)). Whether E-3M2H or apolipoprotein D are involved in these responses have yet to be determined.

Hence, there is an unfulfilled need for the identification of human pheromones and agents, which can augment the pheromone response. Such a response may be useful for the alleviation of anxiety, promotion of beneficial moods, sexual communication, and for the alteration of hypothalamic functions, such as satiety, energy balance, and reproductive biology.

## BRIEF SUMMARY OF THE INVENTION

The present invention provides a novel human pheromone protein, designated "ZHMUP-2," which is a member of the lipocalin family, and structurally related to murine major urinary proteins and porcine sex-specific salivary lipocalin. The present invention also provides ZHMUP-2 variant polypeptides and ZHMUP-2 fusion proteins, as well as nucleic acid molecules encoding such polypeptides and proteins, and methods for using these nucleic acid molecules and amino acid sequences.

## DETAILED DESCRIPTION OF THE INVENTION

### 1. Overview

The present invention provides nucleic acid molecules that encode a new human protein, designated as "ZHMUP-2." An illustrative nucleotide sequence that encodes ZHMUP-2 is provided by SEQ ID NO:1. The encoded polypeptide has the following amino acid sequence: MMLLLLCLGL TLVCAQEEEN NDAVTSNFDL SKISGEWYSV LLASDCREKI EEDGSMRVFV KHIDYLGNSS LTFKLHEIEN GNCTEINLAC KPTEKNAICS TDYNGLNVID ILETDYDNYI YFYKNKIKNG ETFLMLELYV RTPDVSSQLK ERFVKYCEEH GIDKENIFDL TKVDRCLQAR

DEGAA (SEQ ID NO:2). Thus, the ZHMUP-2 gene described herein encodes a polypeptide of 185 amino acids, as shown in SEQ ID NO:2. The predicted signal sequence includes amino acid residues 1 to 15 of SEQ ID NO:2.

Sequence analysis revealed that ZHMUP-2 bears significant homology to  
5 rat major urinary protein. Accordingly, ZHMUP-2 appears to be the orthologous protein of rat major urinary protein, and is considered to be a member of the lipocalin family of proteins.

The *ZHMUP-2* gene, which encodes six exons, resides on human chromosome 6p11.2 – 21, which is a region associated with several disorders, as  
10 described below. A nucleotide sequence that includes the *ZHMUP-2* gene is provided by SEQ ID NO:5. With reference to the amino acid sequence of SEQ ID NO:2, exon 1 encodes amino acid residues 1 to 32, exon 2 encodes amino acid residue 33 to the first two nucleotides of the codon for amino acid residue 78, exon 3 encodes the remainder of amino acid residue 78 to the first two nucleotides of the codon for amino acid residue  
15 103, exon 4 encodes the remainder of amino acid residue 103 to the first nucleotide of the codon for amino acid residue 140, exon 5 encodes the remainder of amino acid residue 140 to the first nucleotide of the codon for amino acid residue 174, and exon 6 encodes the remainder of amino acid residue 174 to amino acid residue 185. Additional features of the nucleotide sequence of SEQ ID NO:5 include an upstream region  
20 (nucleotides 1 to 1751) which contains a TATA box (nucleotides 1720 to 1725) and a potential binding site for transcription factor AP-2 (nucleotides 1687 to 1696). A putative polyadenylation signal resides at nucleotides 5036 to 5042 of SEQ ID NO:5. Table 1 provides further features of this nucleotide sequence.

Table 1

Feature	SEQ ID NO:5	Corresponding region of SEQ ID NO:1
Exon 1	1752-1908 (coding: 1813-1908)	1-96
Intron 1	1909-2259	
Exon 2	2260-2396	97-233
Intron 2	2397-3111	
Exon 3	3112-3185	234-307
Intron 3	3186-4170	
Exon 4	4171-4281	308-418
Intron 4	4282-4622	
Exon 5	4623-4724	419-520
Intron 5	4725-4902	
Exon 6	4903-5072 (coding: 4903-4937)	521-555

A *ZHMUP-2* probe labeled with  $^{32}\text{P}$  by random priming was used to probe mRNA samples isolated from various tissues. The blots were washed with 0.1xSSC containing 0.1% SDS at 65°C. *ZHMUP-2* transcripts of 1.6, 1, and 0.8 kilobases were detected at moderate levels in a variety of tissues, including brain, placenta, lung, thymus, peripheral blood leukocytes, stomach, spinal cord, lymph node, and trachea. Higher *ZHMUP-2* transcript levels were found in heart, liver, skeletal muscle, kidney, pancreas, spleen, prostate, testis, ovaries, small intestine, colon, adrenal gland and bone marrow. *ZHMUP-2* transcripts were not found to be represented by any entries in EST databases, despite the widespread appearance and relatively high levels of *ZHMUP-2* transcripts. This observation suggest that cDNA clones encoding *ZHMUP-2* may be selected against while being propagated through the bacterial host, or that there may be sequences present on the *ZHMUP-2* transcript that inhibit reverse transcription during cDNA synthesis.

As detailed below, the present invention provides isolated polypeptides comprising an amino acid sequence that is at least 70%, at least 80%, or at least 90% identical to a reference amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID NO:2, the amino acid sequence of amino acid residues 46 to 185 of SEQ ID NO:2, the amino acid sequence of amino acid residues 16 to 185 of SEQ ID NO:2, the amino acid sequence of amino acid residues 1 to 32 of SEQ ID NO:2, the amino acid sequence of amino acid residues 16 to 32 of SEQ ID NO:2, the amino

acid sequence of amino acid residues 33 to 77 of SEQ ID NO:2, the amino acid sequence of amino acid residues 79 to 102 of SEQ ID NO:2, the amino acid sequence of amino acid residues 104 to 139 of SEQ ID NO:2, the amino acid sequence of amino acid residues 141 to 173 of SEQ ID NO:2, and the amino acid sequence of amino acid residues 175 to 185 of SEQ ID NO:2. Particular polypeptides specifically bind with an antibody that specifically binds with a polypeptide consisting of the amino acid sequence of SEQ ID NO:2.

Illustrative polypeptides include a polypeptide that comprises the amino acid sequence of SEQ ID NO:2, a polypeptide that consists of the amino acid sequence of amino acid residues 46 to 185 of SEQ ID NO:2, and a polypeptide that consists of any of the following sequences: amino acid residues 1 to 32 of SEQ ID NO:2, amino acid residues 16 to 32 of SEQ ID NO:2, amino acid residues 33 to 77 of SEQ ID NO:2, amino acid residues 79 to 102 of SEQ ID NO:2, amino acid residues 104 to 139 of SEQ ID NO:2, amino acid residues 141 to 173 of SEQ ID NO:2, and amino acid residues 175 to 185 of SEQ ID NO:2.

The present invention further provides polypeptides encoded by at least one *ZHMUP-2* exon. For example, such polypeptides can consist of the following amino acid sequences of SEQ ID NO:2: amino acid residues 1 to 32, amino acid residues 33 to 77, amino acid residues 79 to 102, amino acid residues 104 to 139, amino acid residues 141 to 173, and amino acid residues 175 to amino acid residue 185.

The present invention also includes polypeptides, comprising an amino acid sequence of at least 15, 20, or 30 contiguous amino acids of an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID NO:2, the amino acid sequence of amino acid residues 46 to 185 of SEQ ID NO:2, the amino acid sequence of amino acid residues 16 to 185 of SEQ ID NO:2, the amino acid sequence of amino acid residues 1 to 32 of SEQ ID NO:2, the amino acid sequence of amino acid residues 16 to 32 of SEQ ID NO:2, the amino acid sequence of amino acid residues 33 to 77 of SEQ ID NO:2, the amino acid sequence of amino acid residues 79 to 102 of SEQ ID NO:2, the amino acid sequence of amino acid residues 104 to 139 of SEQ ID NO:2, the amino acid sequence of amino acid residues 141 to 173 of SEQ ID NO:2, and the amino acid sequence of amino acid residues 175 to 185 of SEQ ID NO:2.

The present invention further provides antibodies and antibody fragments that specifically bind with such polypeptides. Exemplary antibodies include polyclonal antibodies, murine monoclonal antibodies, humanized antibodies derived from murine monoclonal antibodies, and human monoclonal antibodies. Illustrative antibody fragments include  $F(ab')_2$ ,  $F(ab)_2$ ,  $Fab'$ ,  $Fab$ ,  $Fv$ ,  $scFv$ , and minimal recognition units. The present invention also includes anti-idiotypic antibodies that specifically bind with



such antibodies or antibody fragments. The present invention further includes compositions comprising a carrier and a peptide, polypeptide, antibody, or anti-idiotypic antibody described herein.

The present invention also provides isolated nucleic acid molecules that  
5 encode a ZHMUP-2 polypeptide, wherein the nucleic acid molecule is selected from the group consisting of (a) a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:3, (b) a nucleic acid molecule encoding the amino acid sequence of SEQ ID NO:2, and (c) a nucleic acid molecule that remains hybridized following stringent wash  
10 conditions to a nucleic acid molecule consisting of a nucleotide sequence, or the complement of a nucleotide sequence, selected from the group consisting of: nucleotides 1 to 555 of SEQ ID NO:1, and nucleotides 46 to 555 of SEQ ID NO:1.

Illustrative nucleic acid molecules include those in which any difference between the amino acid sequence encoded by the nucleic acid molecule and the corresponding amino acid sequence of SEQ ID NO:2 is due to a conservative amino acid  
15 substitution. The present invention further contemplates isolated nucleic acid molecules that comprise the nucleotide sequence of SEQ ID NO:1, or the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1.

The present invention also provides nucleic acid molecules that consist of the nucleotide sequence of a ZHMUP-2 exon or intron. The nucleotide sequences of  
20 these exons and introns are identified in Table 1.

The present invention also includes vectors and expression vectors comprising such nucleic acid molecules. Such expression vectors may comprise a transcription promoter, and a transcription terminator, wherein the promoter is operably linked with the nucleic acid molecule, and wherein the nucleic acid molecule is operably  
25 linked with the transcription terminator. The present invention further includes recombinant host cells comprising these vectors and expression vectors. Illustrative host cells include bacterial, yeast, fungal, insect, avian, mammalian, and plant cells. Recombinant host cells comprising such expression vectors can be used to produce ZHMUP-2 polypeptides by culturing such recombinant host cells that comprise the  
30 expression vector and that produce the ZHMUP-2 protein, and, optionally, isolating the ZHMUP-2 protein from the cultured recombinant host cells. The present invention further includes products made by these processes.

The present invention also contemplates methods for detecting the presence of ZHMUP-2 RNA in a biological sample, comprising the steps of (a)  
35 contacting a ZHMUP-2 nucleic acid probe under hybridizing conditions with either (i) test RNA molecules isolated from the biological sample, or (ii) nucleic acid molecules synthesized from the isolated RNA molecules, wherein the probe has a nucleotide

sequence comprising a portion of the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1, or its complement, and (b) detecting the formation of hybrids of the nucleic acid probe and either the test RNA molecules or the synthesized nucleic acid molecules, wherein the presence of the hybrids indicates the presence of *ZHMUP-2* RNA in the biological sample. An example of a biological sample is a human biological sample, such as a biopsy or autopsy specimen.

The present invention further provides methods for detecting the presence of *ZHMUP-2* polypeptide in a biological sample, comprising the steps of: (a) contacting the biological sample with an antibody or an antibody fragment that specifically binds with a polypeptide having the amino acid sequence of SEQ ID NO:2, wherein the contacting is performed under conditions that allow the binding of the antibody or antibody fragment to the biological sample, and (b) detecting any of the bound antibody or bound antibody fragment. Such an antibody or antibody fragment may further comprise a detectable label selected from the group consisting of radioisotope, fluorescent label, chemiluminescent label, enzyme label, bioluminescent label, and colloidal gold. An exemplary biological sample is a human biological sample.

The present invention also provides kits for performing these detection methods. For example, a kit for detection of *ZHMUP-2* gene expression may comprise a container that comprises a nucleic acid molecule, wherein the nucleic acid molecule is selected from the group consisting of (a) a nucleic acid molecule comprising the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1, (b) a nucleic acid molecule comprising the complement of the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1, (c) a nucleic acid molecule that is a fragment of (a) consisting of at least eight nucleotides, and (d) a nucleic acid molecule that is a fragment of (b) consisting of at least eight nucleotides. Such a kit may also comprise a second container that comprises one or more reagents capable of indicating the presence of the nucleic acid molecule. On the other hand, a kit for detection of *ZHMUP-2* protein may comprise a container that comprises an antibody, or an antibody fragment, that specifically binds with a polypeptide having the amino acid sequence of SEQ ID NO:2.

The present invention further provides fusion proteins a *ZHMUP-2* polypeptide and an immunoglobulin moiety. In such fusion proteins, the immunoglobulin moiety may be an immunoglobulin heavy chain constant region, such as a human  $F_c$  fragment. The present invention further includes isolated nucleic acid molecules that encode such fusion proteins.

The present invention also includes methods for detecting a chromosome 6p11.2 – 21 abnormality in a subject by (a) amplifying, from genomic DNA isolated from a biological sample of the subject, nucleic acid molecules that either (i) comprise a

nucleotide sequence that encodes at least one of *ZHMUP-2* exons 1 to 6, or that (ii) comprise a nucleotide sequence that is the complement of (i), and (b) detecting a mutation in the amplified nucleic acid molecules, wherein the presence of a mutation indicates a chromosome 6p11.2 – 21 abnormality. In variations of these methods, the detecting step is performed by comparing the nucleotide sequence of the amplified nucleic acid molecules to the nucleotide sequence of SEQ ID NOs:1 or 5.

Similarly, the present invention provides methods for detecting a chromosome 6p11.2 – 21 abnormality in a subject comprising: (a) amplifying, from genomic DNA isolated from a biological sample of the subject, a segment of the *ZHMUP-2* gene that comprises either the nucleotide sequence of any one of introns 1 to 5, or the complementary nucleotide sequence of any one of introns 1 to 5, and (b) detecting a mutation in the amplified nucleic acid molecules, wherein the presence of a mutation indicates a chromosome 6p11.2 – 21 abnormality. In variations of these methods, the detecting step is performed by binding the amplified *ZHMUP-2* gene segments to a membrane, and contacting the membrane with a nucleic acid probe under hybridizing conditions of high stringency, wherein the absence of hybrids indicates metabolic disease or susceptibility to a metabolic disease, or a mutation in chromosome 6p11.2 – 21. As an illustration, a nucleic acid probe can comprise the nucleotide sequence (or the complementary nucleotide sequence) of any one of introns 1 to 5.

Examples of mutations or alterations of the *ZHMUP-2* gene or its gene products include point mutations, deletions, insertions, and rearrangements. Another example of a *ZHMUP-2* gene mutation is aneuploidy.

These and other aspects of the invention will become evident upon reference to the following detailed description. In addition, various references are identified below.

## 2. Definitions

In the description that follows, a number of terms are used extensively. The following definitions are provided to facilitate understanding of the invention.

As used herein, “nucleic acid” or “nucleic acid molecule” refers to polynucleotides, such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), oligonucleotides, fragments generated by the polymerase chain reaction (PCR), and fragments generated by any of ligation, scission, endonuclease action, and exonuclease action. Nucleic acid molecules can be composed of monomers that are naturally-

occurring nucleotides (such as DNA and RNA), or analogs of naturally-occurring nucleotides (*e.g.*,  $\alpha$ -enantiomeric forms of naturally-occurring nucleotides), or a combination of both. Modified nucleotides can have alterations in sugar moieties and/or in pyrimidine or purine base moieties. Sugar modifications include, for example, replacement of one or more hydroxyl groups with halogens, alkyl groups, amines, and azido groups, or sugars can be functionalized as ethers or esters. Moreover, the entire sugar moiety can be replaced with sterically and electronically similar structures, such as aza-sugars and carbocyclic sugar analogs. Examples of modifications in a base moiety include alkylated purines and pyrimidines, acylated purines or pyrimidines, or other well-known heterocyclic substitutes. Nucleic acid monomers can be linked by phosphodiester bonds or analogs of such linkages. Analogs of phosphodiester linkages include phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoranilidate, phosphoramidate, and the like. The term “nucleic acid molecule” also includes so-called “peptide nucleic acids,” which comprise naturally-occurring or modified nucleic acid bases attached to a polyamide backbone. Nucleic acids can be either single stranded or double stranded.

The term “complement of a nucleic acid molecule” refers to a nucleic acid molecule having a complementary nucleotide sequence and reverse orientation as compared to a reference nucleotide sequence. For example, the sequence 5' ATGCACGGG 3' is complementary to 5' CCCGTGCAT 3'.

The term “contig” denotes a nucleic acid molecule that has a contiguous stretch of identical or complementary sequence to another nucleic acid molecule. Contiguous sequences are said to “overlap” a given stretch of a nucleic acid molecule either in their entirety or along a partial stretch of the nucleic acid molecule.

The term “degenerate nucleotide sequence” denotes a sequence of nucleotides that includes one or more degenerate codons as compared to a reference nucleic acid molecule that encodes a polypeptide. Degenerate codons contain different triplets of nucleotides, but encode the same amino acid residue (*i.e.*, GAU and GAC triplets each encode Asp).

The term “structural gene” refers to a nucleic acid molecule that is transcribed into messenger RNA (mRNA), which is then translated into a sequence of amino acids characteristic of a specific polypeptide.

An “isolated nucleic acid molecule” is a nucleic acid molecule that is not integrated in the genomic DNA of an organism. For example, a DNA molecule that encodes a growth factor that has been separated from the genomic DNA of a cell is an isolated DNA molecule. Another example of an isolated nucleic acid molecule is a chemically-synthesized nucleic acid molecule that is not integrated in the genome of an

organism. A nucleic acid molecule that has been isolated from a particular species is smaller than the complete DNA molecule of a chromosome from that species.

A "nucleic acid molecule construct" is a nucleic acid molecule, either single- or double-stranded, that has been modified through human intervention to contain segments of nucleic acid combined and juxtaposed in an arrangement not existing in nature.

"Linear DNA" denotes non-circular DNA molecules having free 5' and 3' ends. Linear DNA can be prepared from closed circular DNA molecules, such as plasmids, by enzymatic digestion or physical disruption.

"Complementary DNA (cDNA)" is a single-stranded DNA molecule that is formed from an mRNA template by the enzyme reverse transcriptase. Typically, a primer complementary to portions of mRNA is employed for the initiation of reverse transcription. Those skilled in the art also use the term "cDNA" to refer to a double-stranded DNA molecule consisting of such a single-stranded DNA molecule and its complementary DNA strand. The term "cDNA" also refers to a clone of a cDNA molecule synthesized from an RNA template.

A "promoter" is a nucleotide sequence that directs the transcription of a structural gene. Typically, a promoter is located in the 5' non-coding region of a gene, proximal to the transcriptional start site of a structural gene. Sequence elements within promoters that function in the initiation of transcription are often characterized by consensus nucleotide sequences. These promoter elements include RNA polymerase binding sites, TATA sequences, CAAT sequences, differentiation-specific elements (DSEs; McGehee *et al.*, *Mol. Endocrinol.* 7:551 (1993)), cyclic AMP response elements (CREs), serum response elements (SREs; Treisman, *Seminars in Cancer Biol.* 1:47 (1990)), glucocorticoid response elements (GREs), and binding sites for other transcription factors, such as CRE/ATF (O'Reilly *et al.*, *J. Biol. Chem.* 267:19938 (1992)), AP2 (Ye *et al.*, *J. Biol. Chem.* 269:25728 (1994)), SP1, cAMP response element binding protein (CREB; Loeken, *Gene Expr.* 3:253 (1993)) and octamer factors (see, in general, Watson *et al.*, eds., *Molecular Biology of the Gene*, 4th ed. (The Benjamin/Cummings Publishing Company, Inc. 1987), and Lemaigre and Rousseau, *Biochem. J.* 303:1 (1994)). If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter. Repressible promoters are also known.

A "core promoter" contains essential nucleotide sequences for promoter function, including the TATA box and start of transcription. By this definition, a core

promoter may or may not have detectable activity in the absence of specific sequences that may enhance the activity or confer tissue specific activity.

A "regulatory element" is a nucleotide sequence that modulates the activity of a core promoter. For example, a regulatory element may contain a nucleotide  
5 sequence that binds with cellular factors enabling transcription exclusively or preferentially in particular cells, tissues, or organelles. These types of regulatory elements are normally associated with genes that are expressed in a "cell-specific," "tissue-specific," or "organelle-specific" manner.

An "enhancer" is a type of regulatory element that can increase the  
10 efficiency of transcription, regardless of the distance or orientation of the enhancer relative to the start site of transcription.

"Heterologous DNA" refers to a DNA molecule, or a population of DNA molecules, that does not exist naturally within a given host cell. DNA molecules heterologous to a particular host cell may contain DNA derived from the host cell  
15 species (*i.e.*, endogenous DNA) so long as that host DNA is combined with non-host DNA (*i.e.*, exogenous DNA). For example, a DNA molecule containing a non-host DNA segment encoding a polypeptide operably linked to a host DNA segment comprising a transcription promoter is considered to be a heterologous DNA molecule. Conversely, a heterologous DNA molecule can comprise an endogenous gene operably  
20 linked with an exogenous promoter. As another illustration, a DNA molecule comprising a gene derived from a wild-type cell is considered to be heterologous DNA if that DNA molecule is introduced into a mutant cell that lacks the wild-type gene.

A "polypeptide" is a polymer of amino acid residues joined by peptide bonds, whether produced naturally or synthetically. Polypeptides of less than about 10  
25 amino acid residues are commonly referred to as "peptides."

A "protein" is a macromolecule comprising one or more polypeptide chains. A protein may also comprise non-peptidic components, such as carbohydrate groups. Carbohydrates and other non-peptidic substituents may be added to a protein by the cell in which the protein is produced, and will vary with the type of cell. Proteins  
30 are defined herein in terms of their amino acid backbone structures; substituents such as carbohydrate groups are generally not specified, but may be present nonetheless.

A peptide or polypeptide encoded by a non-host DNA molecule is a "heterologous" peptide or polypeptide.

An "integrated genetic element" is a segment of DNA that has been  
35 incorporated into a chromosome of a host cell after that element is introduced into the cell through human manipulation. Within the present invention, integrated genetic elements are most commonly derived from linearized plasmids that are introduced into

the cells by electroporation or other techniques. Integrated genetic elements are passed from the original host cell to its progeny.

A "cloning vector" is a nucleic acid molecule, such as a plasmid, cosmid, or bacteriophage, that has the capability of replicating autonomously in a host cell. Cloning vectors typically contain one or a small number of restriction endonuclease recognition sites that allow insertion of a nucleic acid molecule in a determinable fashion without loss of an essential biological function of the vector, as well as nucleotide sequences encoding a marker gene that is suitable for use in the identification and selection of cells transformed with the cloning vector. Marker genes typically include genes that provide tetracycline resistance or ampicillin resistance.

An "expression vector" is a nucleic acid molecule encoding a gene that is expressed in a host cell. Typically, an expression vector comprises a transcription promoter, a gene, and a transcription terminator. Gene expression is usually placed under the control of a promoter, and such a gene is said to be "operably linked to" the promoter. Similarly, a regulatory element and a core promoter are operably linked if the regulatory element modulates the activity of the core promoter.

A "recombinant host" is a cell that contains a heterologous nucleic acid molecule, such as a cloning vector or expression vector. In the present context, an example of a recombinant host is a cell that produces ZHMUP-2 from an expression vector. In contrast, ZHMUP-2 can be produced by a cell that is a "natural source" of ZHMUP-2, and that lacks an expression vector.

"Integrative transformants" are recombinant host cells, in which heterologous DNA has become integrated into the genomic DNA of the cells.

A "fusion protein" is a hybrid protein expressed by a nucleic acid molecule comprising nucleotide sequences of at least two genes. For example, a fusion protein can comprise at least part of a ZHMUP-2 polypeptide fused with a polypeptide that binds an affinity matrix. Such a fusion protein provides a means to isolate large quantities of ZHMUP-2 using affinity chromatography.

The term "receptor" denotes a cell-associated protein that binds to a bioactive molecule termed a "ligand." This interaction mediates the effect of the ligand on the cell. Receptors can be membrane bound, cytosolic or nuclear; monomeric (*e.g.*, thyroid stimulating hormone receptor, beta-adrenergic receptor) or multimeric (*e.g.*, PDGF receptor, growth hormone receptor, IL-3 receptor, GM-CSF receptor, G-CSF receptor, erythropoietin receptor and IL-6 receptor). Membrane-bound receptors are characterized by a multi-domain structure comprising an extracellular ligand-binding domain and an intracellular effector domain that is typically involved in signal transduction. In certain membrane-bound receptors, the extracellular ligand-binding

domain and the intracellular effector domain are located in separate polypeptides that comprise the complete functional receptor.

In general, the binding of ligand to receptor results in a conformational change in the receptor that causes an interaction between the effector domain and other molecule(s) in the cell, which in turn leads to an alteration in the metabolism of the cell. Metabolic events that are often linked to receptor-ligand interactions include gene transcription, phosphorylation, dephosphorylation, increases in cyclic AMP production, mobilization of cellular calcium, mobilization of membrane lipids, cell adhesion, hydrolysis of inositol lipids and hydrolysis of phospholipids.

The term "secretory signal sequence" denotes a nucleotide sequence that encodes a peptide (a "secretory peptide") that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in which it is synthesized. The larger polypeptide is commonly cleaved to remove the secretory peptide during transit through the secretory pathway.

An "isolated polypeptide" is a polypeptide that is essentially free from contaminating cellular components, such as carbohydrate, lipid, or other proteinaceous impurities associated with the polypeptide in nature. Typically, a preparation of isolated polypeptide contains the polypeptide in a highly purified form, *i.e.*, at least about 80% pure, at least about 90% pure, at least about 95% pure, greater than 95% pure, or greater than 99% pure. One way to show that a particular protein preparation contains an isolated polypeptide is by the appearance of a single band following sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis of the protein preparation and Coomassie Brilliant Blue staining of the gel. However, the term "isolated" does not exclude the presence of the same polypeptide in alternative physical forms, such as dimers or alternatively glycosylated or derivatized forms.

The terms "amino-terminal" and "carboxyl-terminal" are used herein to denote positions within polypeptides. Where the context allows, these terms are used with reference to a particular sequence or portion of a polypeptide to denote proximity or relative position. For example, a certain sequence positioned carboxyl-terminal to a reference sequence within a polypeptide is located proximal to the carboxyl terminus of the reference sequence, but is not necessarily at the carboxyl terminus of the complete polypeptide.

The term "expression" refers to the biosynthesis of a gene product. For example, in the case of a structural gene, expression involves transcription of the structural gene into mRNA and the translation of mRNA into one or more polypeptides.

The term "splice variant" is used herein to denote alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative



splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a polypeptide encoded by  
5 a splice variant of an mRNA transcribed from a gene.

As used herein, the term "immunomodulator" includes cytokines, stem cell growth factors, lymphotoxins, co-stimulatory molecules, hematopoietic factors, and synthetic analogs of these molecules.

The term "complement/anti-complement pair" denotes non-identical  
10 moieties that form a non-covalently associated, stable pair under appropriate conditions. For instance, biotin and avidin (or streptavidin) are prototypical members of a complement/anti-complement pair. Other exemplary complement/anti-complement pairs include receptor/ligand pairs, antibody/antigen (or hapten or epitope) pairs, sense/antisense polynucleotide pairs, and the like. Where subsequent dissociation of the  
15 complement/anti-complement pair is desirable, the complement/anti-complement pair preferably has a binding affinity of less than  $10^9 \text{ M}^{-1}$ .

An "anti-idiotypic antibody" is an antibody that binds with the variable region domain of an immunoglobulin. In the present context, an anti-idiotypic antibody binds with the variable region of an anti-ZHMUP-2 antibody, and thus, an anti-idiotypic  
20 antibody mimics an epitope of ZHMUP-2.

An "antibody fragment" is a portion of an antibody such as  $\text{F(ab')}_2$ ,  $\text{F(ab)}_2$ ,  $\text{Fab'}$ ,  $\text{Fab}$ , and the like. Regardless of structure, an antibody fragment binds with the same antigen that is recognized by the intact antibody. For example, an anti-ZHMUP-2 monoclonal antibody fragment binds with an epitope of ZHMUP-2.

The term "antibody fragment" also includes a synthetic or a genetically  
25 engineered polypeptide that binds to a specific antigen, such as polypeptides consisting of the light chain variable region, "Fv" fragments consisting of the variable regions of the heavy and light chains, recombinant single chain polypeptide molecules in which light and heavy variable regions are connected by a peptide linker ("scFv proteins"), and minimal  
30 recognition units consisting of the amino acid residues that mimic the hypervariable region.

A "chimeric antibody" is a recombinant protein that contains the variable domains and complementary determining regions derived from a rodent antibody, while the remainder of the antibody molecule is derived from a human antibody.

"Humanized antibodies" are recombinant proteins in which murine  
35 complementarity determining regions of a monoclonal antibody have been transferred from heavy and light variable chains of the murine immunoglobulin into a human variable domain.

As used herein, a “therapeutic agent” is a molecule or atom which is conjugated to an antibody moiety to produce a conjugate which is useful for therapy. Examples of therapeutic agents include drugs, toxins, immunomodulators, chelators, boron compounds, photoactive agents or dyes, and radioisotopes.

5 A “detectable label” is a molecule or atom which can be conjugated to an antibody moiety to produce a molecule useful for diagnosis. Examples of detectable labels include chelators, photoactive agents, radioisotopes, fluorescent agents, paramagnetic ions, or other marker moieties.

The term “affinity tag” is used herein to denote a polypeptide segment  
10 that can be attached to a second polypeptide to provide for purification or detection of the second polypeptide or provide sites for attachment of the second polypeptide to a substrate. In principal, any peptide or protein for which an antibody or other specific binding agent is available can be used as an affinity tag. Affinity tags include a poly-histidine tract, protein A (Nilsson *et al.*, *EMBO J.* 4:1075 (1985); Nilsson *et al.*, *Methods*  
15 *Enzymol.* 198:3 (1991)), glutathione S transferase (Smith and Johnson, *Gene* 67:31 (1988)), Glu-Glu affinity tag (Grussenmeyer *et al.*, *Proc. Natl. Acad. Sci. USA* 82:7952 (1985)), substance P, FLAG peptide (Hopp *et al.*, *Biotechnology* 6:1204 (1988)), streptavidin binding peptide, or other antigenic epitope or binding domain. See, in general, Ford *et al.*, *Protein Expression and Purification* 2:95 (1991). Nucleic acid  
20 molecules encoding affinity tags are available from commercial suppliers (*e.g.*, Pharmacia Biotech, Piscataway, NJ).

A “naked antibody” is an entire antibody, as opposed to an antibody fragment, which is not conjugated with a therapeutic agent. Naked antibodies include both polyclonal and monoclonal antibodies, as well as certain recombinant antibodies,  
25 such as chimeric and humanized antibodies.

As used herein, the term “antibody component” includes both an entire antibody and an antibody fragment.

An “immunoconjugate” is a conjugate of an antibody component with a therapeutic agent or a detectable label.

30 As used herein, the term “antibody fusion protein” refers to a recombinant molecule that comprises an antibody component and a therapeutic agent. Examples of therapeutic agents suitable for such fusion proteins include immunomodulators (“antibody-immunomodulator fusion protein”) and toxins (“antibody-toxin fusion protein”).

35 A “target polypeptide” or a “target peptide” is an amino acid sequence that comprises at least one epitope, and that is expressed on a target cell, such as a tumor cell, or a cell that carries an infectious agent antigen. T cells recognize peptide epitopes

presented by a major histocompatibility complex molecule to a target polypeptide or target peptide and typically lyse the target cell or recruit other immune cells to the site of the target cell, thereby killing the target cell.

An "antigenic peptide" is a peptide which will bind a major  
5 histocompatibility complex molecule to form an MHC-peptide complex which is recognized by a T cell, thereby inducing a cytotoxic lymphocyte response upon presentation to the T cell. Thus, antigenic peptides are capable of binding to an appropriate major histocompatibility complex molecule and inducing a cytotoxic T cells response, such as cell lysis or specific cytokine release against the target cell which binds  
10 or expresses the antigen. The antigenic peptide can be bound in the context of a class I or class II major histocompatibility complex molecule, on an antigen presenting cell or on a target cell.

In eukaryotes, RNA polymerase II catalyzes the transcription of a structural gene to produce mRNA. A nucleic acid molecule can be designed to contain  
15 an RNA polymerase II template in which the RNA transcript has a sequence that is complementary to that of a specific mRNA. The RNA transcript is termed an "anti-sense RNA" and a nucleic acid molecule that encodes the anti-sense RNA is termed an "anti-sense gene." Anti-sense RNA molecules are capable of binding to mRNA molecules, resulting in an inhibition of mRNA translation.

20 An "anti-sense oligonucleotide specific for *ZHMUP-2*" or an "*ZHMUP-2* anti-sense oligonucleotide" is an oligonucleotide having a sequence (a) capable of forming a stable triplex with a portion of the *ZHMUP-2* gene, or (b) capable of forming a stable duplex with a portion of an mRNA transcript of the *ZHMUP-2* gene.

A "ribozyme" is a nucleic acid molecule that contains a catalytic center.  
25 The term includes RNA enzymes, self-splicing RNAs, self-cleaving RNAs, and nucleic acid molecules that perform these catalytic functions. A nucleic acid molecule that encodes a ribozyme is termed a "ribozyme gene."

An "external guide sequence" is a nucleic acid molecule that directs the endogenous ribozyme, RNase P, to a particular species of intracellular mRNA, resulting  
30 in the cleavage of the mRNA by RNase P. A nucleic acid molecule that encodes an external guide sequence is termed an "external guide sequence gene."

The term "variant *ZHMUP-2* gene" refers to nucleic acid molecules that encode a polypeptide having an amino acid sequence that is a modification of SEQ ID NO:2. Such variants include naturally-occurring polymorphisms of *ZHMUP-2* genes, as  
35 well as synthetic genes that contain conservative amino acid substitutions of the amino acid sequence of SEQ ID NO:2. Additional variant forms of *ZHMUP-2* genes are nucleic acid molecules that contain insertions or deletions of the nucleotide sequences

described herein. A variant *ZHMUP-2* gene can be identified by determining whether the gene hybridizes with a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, or its complement, under stringent conditions.

Alternatively, variant *ZHMUP-2* genes can be identified by sequence  
5 comparison. Two amino acid sequences have "100% amino acid sequence identity" if the amino acid residues of the two amino acid sequences are the same when aligned for maximal correspondence. Similarly, two nucleotide sequences have "100% nucleotide sequence identity" if the nucleotide residues of the two nucleotide sequences are the same when aligned for maximal correspondence. Sequence comparisons can be  
10 performed using standard software programs such as those included in the LASERGENE bioinformatics computing suite, which is produced by DNASTAR (Madison, Wisconsin). Other methods for comparing two nucleotide or amino acid sequences by determining optimal alignment are well-known to those of skill in the art (see, for example, Peruski and Peruski, *The Internet and the New Biology: Tools for Genomic and*  
15 *Molecular Research* (ASM Press, Inc. 1997), Wu *et al.* (eds.), "Information Superhighway and Computer Databases of Nucleic Acids and Proteins," in *Methods in Gene Biotechnology*, pages 123-151 (CRC Press, Inc. 1997), and Bishop (ed.), *Guide to Human Genome Computing*, 2nd Edition (Academic Press, Inc. 1998)). Particular methods for determining sequence identity are described below.

20 Regardless of the particular method used to identify a variant *ZHMUP-2* gene or variant *ZHMUP-2* polypeptide, a variant gene or polypeptide encoded by a variant gene may be characterized by the ability to bind specifically to an anti-*ZHMUP-2* antibody.

The term "allelic variant" is used herein to denote any of two or more  
25 alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene.

30 The term "ortholog" denotes a polypeptide or protein obtained from one species that is the functional counterpart of a polypeptide or protein from a different species. Sequence differences among orthologs are the result of speciation.

"Paralogs" are distinct but structurally related proteins made by an organism. Paralogs are believed to arise through gene duplication. For example,  $\alpha$ -globin,  $\beta$ -globin, and myoglobin are paralogs of each other.  
35

The present invention includes functional fragments of *ZHMUP-2* genes. Within the context of this invention, a "functional fragment" of a *ZHMUP-2* gene refers

to a nucleic acid molecule that encodes a portion of a ZHMUP-2 polypeptide which specifically binds with an anti-ZHMUP-2 antibody. For example, a functional fragment of a ZHMUP-2 gene described herein comprises a portion of the nucleotide sequence of SEQ ID NO:1, and encodes a polypeptide that specifically binds with an anti-ZHMUP-2 antibody.

Due to the imprecision of standard analytical methods, molecular weights and lengths of polymers are understood to be approximate values. When such a value is expressed as "about" X or "approximately" X, the stated value of X will be understood to be accurate to  $\pm 10\%$ .

### 3. *Production of Nucleic Acid Molecules Encoding ZHMUP-2*

Nucleic acid molecules encoding ZHMUP-2 can be obtained by screening a human cDNA or genomic library using polynucleotide probes based upon SEQ ID NO:1. These techniques are standard and well-established.

As an illustration, a nucleic acid molecule that encodes human ZHMUP-2 can be isolated from a human cDNA library. In this case, the first step would be to prepare the cDNA library by isolating RNA from tissue, using methods well-known to those of skill in the art. In general, RNA isolation techniques must provide a method for breaking cells, a means of inhibiting RNase-directed degradation of RNA, and a method of separating RNA from DNA, protein, and polysaccharide contaminants. For example, total RNA can be isolated by freezing tissue in liquid nitrogen, grinding the frozen tissue with a mortar and pestle to lyse the cells, extracting the ground tissue with a solution of phenol/chloroform to remove proteins, and separating RNA from the remaining impurities by selective precipitation with lithium chloride (see, for example, Ausubel *et al.* (eds.), *Short Protocols in Molecular Biology*, 3<sup>rd</sup> Edition, pages 4-1 to 4-6 (John Wiley & Sons 1995) ["Ausubel (1995)"]; Wu *et al.*, *Methods in Gene Biotechnology*, pages 33-41 (CRC Press, Inc. 1997) ["Wu (1997)"]).

Alternatively, total RNA can be isolated from tissue by extracting ground tissue with guanidinium isothiocyanate, extracting with organic solvents, and separating RNA from contaminants using differential centrifugation (see, for example, Chirgwin *et al.*, *Biochemistry* 18:52 (1979); Ausubel (1995) at pages 4-1 to 4-6; Wu (1997) at pages 33-41).

In order to construct a cDNA library, poly(A)<sup>+</sup> RNA must be isolated from a total RNA preparation. Poly(A)<sup>+</sup> RNA can be isolated from total RNA using the standard technique of oligo(dT)-cellulose chromatography (see, for example, Aviv and Leder, *Proc. Nat'l Acad. Sci. USA* 69:1408 (1972); Ausubel (1995) at pages 4-11 to 4-12).

Double-stranded cDNA molecules are synthesized from poly(A)<sup>+</sup> RNA using techniques well-known to those in the art. (see, for example, Wu (1997) at pages 41-46). Moreover, commercially available kits can be used to synthesize double-stranded cDNA molecules. For example, such kits are available from Life Technologies, Inc. (Gaithersburg, MD), CLONTECH Laboratories, Inc. (Palo Alto, CA), Promega Corporation (Madison, WI) and STRATAGENE (La Jolla, CA).

Various cloning vectors are appropriate for the construction of a cDNA library. For example, a cDNA library can be prepared in a vector derived from bacteriophage, such as a  $\lambda$ gt10 vector. See, for example, Huynh *et al.*, "Constructing and Screening cDNA Libraries in  $\lambda$ gt10 and  $\lambda$ gt11," in *DNA Cloning: A Practical Approach Vol. I*, Glover (ed.), page 49 (IRL Press, 1985); Wu (1997) at pages 47-52.

Alternatively, double-stranded cDNA molecules can be inserted into a plasmid vector, such as a PBLUESCRIPT vector (STRATAGENE; La Jolla, CA), a LAMDA GEM-4 (Promega Corp.) or other commercially available vectors. Suitable cloning vectors also can be obtained from the American Type Culture Collection (Manassas, VA).

To amplify the cloned cDNA molecules, the cDNA library is inserted into a prokaryotic host, using standard techniques. For example, a cDNA library can be introduced into competent *E. coli* DH5 cells, which can be obtained, for example, from Life Technologies, Inc. (Gaithersburg, MD).

A human genomic library can be prepared by means well-known in the art (see, for example, Ausubel (1995) at pages 5-1 to 5-6; Wu (1997) at pages 307-327). Genomic DNA can be isolated by lysing tissue with the detergent Sarkosyl, digesting the lysate with proteinase K, clearing insoluble debris from the lysate by centrifugation, precipitating nucleic acid from the lysate using isopropanol, and purifying resuspended DNA on a cesium chloride density gradient.

DNA fragments that are suitable for the production of a genomic library can be obtained by the random shearing of genomic DNA or by the partial digestion of genomic DNA with restriction endonucleases. Genomic DNA fragments can be inserted into a vector, such as a bacteriophage or cosmid vector, in accordance with conventional techniques, such as the use of restriction enzyme digestion to provide appropriate termini, the use of alkaline phosphatase treatment to avoid undesirable joining of DNA molecules, and ligation with appropriate ligases. Techniques for such manipulation are well-known in the art (see, for example, Ausubel (1995) at pages 5-1 to 5-6; Wu (1997) at pages 307-327).

Nucleic acid molecules that encode a human *ZHMUP-2* gene can also be obtained using the polymerase chain reaction (PCR) with oligonucleotide primers having nucleotide sequences that are based upon the nucleotide sequences of the human

*ZHMUP-2* gene, as described herein. General methods for screening libraries with PCR are provided by, for example, Yu *et al.*, "Use of the Polymerase Chain Reaction to Screen Phage Libraries," in *Methods in Molecular Biology, Vol. 15: PCR Protocols: Current Methods and Applications*, White (ed.), pages 211-215 (Humana Press, Inc. 5 1993). Moreover, techniques for using PCR to isolate related genes are described by, for example, Preston, "Use of Degenerate Oligonucleotide Primers and the Polymerase Chain Reaction to Clone Gene Family Members," in *Methods in Molecular Biology, Vol. 15: PCR Protocols: Current Methods and Applications*, White (ed.), pages 317-337 (Humana Press, Inc. 1993).

10 Alternatively, human genomic libraries can be obtained from commercial sources such as Research Genetics (Huntsville, AL) and the American Type Culture Collection (Manassas, VA).

A library containing cDNA or genomic clones can be screened with one or more polynucleotide probes based upon SEQ ID NO:1, using standard methods (see, for 15 example, Ausubel (1995) at pages 6-1 to 6-11).

Anti-*ZHMUP-2* antibodies, produced as described below, can also be used to isolate DNA sequences that encode human *ZHMUP-2* genes from cDNA libraries. For example, the antibodies can be used to screen  $\lambda$ gt11 expression libraries, or the antibodies can be used for immunoscreening following hybrid selection and 20 translation (see, for example, Ausubel (1995) at pages 6-12 to 6-16; Margolis *et al.*, "Screening  $\lambda$  expression libraries with antibody and protein probes," in *DNA Cloning 2: Expression Systems, 2nd Edition*, Glover *et al.* (eds.), pages 1-14 (Oxford University Press 1995)).

As an alternative, a *ZHMUP-2* gene can be obtained by synthesizing 25 nucleic acid molecules using mutually priming long oligonucleotides and the nucleotide sequences described herein (see, for example, Ausubel (1995) at pages 8-8 to 8-9). Established techniques using the polymerase chain reaction provide the ability to synthesize DNA molecules at least two kilobases in length (Adang *et al.*, *Plant Molec. Biol.* 21:1131 (1993), Bambot *et al.*, *PCR Methods and Applications* 2:266 (1993), 30 Dillon *et al.*, "Use of the Polymerase Chain Reaction for the Rapid Construction of Synthetic Genes," in *Methods in Molecular Biology, Vol. 15: PCR Protocols: Current Methods and Applications*, White (ed.), pages 263-268, (Humana Press, Inc. 1993), and Holowachuk *et al.*, *PCR Methods Appl.* 4:299 (1995)).

The nucleic acid molecules of the present invention can also be 35 synthesized with "gene machines" using protocols such as the phosphoramidite method. If chemically-synthesized double stranded DNA is required for an application such as the synthesis of a gene or a gene fragment, then each complementary strand is made

separately. The production of short genes (60 to 80 base pairs) is technically straightforward and can be accomplished by synthesizing the complementary strands and then annealing them. For the production of longer genes (>300 base pairs), however, special strategies may be required, because the coupling efficiency of each cycle during chemical DNA synthesis is seldom 100%. To overcome this problem, synthetic genes (double-stranded) are assembled in modular form from single-stranded fragments that are from 20 to 100 nucleotides in length. For reviews on polynucleotide synthesis, see, for example, Glick and Pasternak, *Molecular Biotechnology, Principles and Applications of Recombinant DNA* (ASM Press 1994), Itakura *et al.*, *Annu. Rev. Biochem.* 53:323 (1984), and Climie *et al.*, *Proc. Nat'l Acad. Sci. USA* 87:633 (1990).

The sequence of a *ZHMUP-2* cDNA or *ZHMUP-2* genomic fragment can be determined using standard methods. *ZHMUP-2* polynucleotide sequences disclosed herein can also be used as probes or primers to clone 5' noncoding regions of a *ZHMUP-2* gene. Promoter elements from a *ZHMUP-2* gene can be used to direct the expression of heterologous genes in, for example, tissue of transgenic animals, or in patients undergoing gene therapy. Such a promoter element can be provided by a fragment consisting of 50, 100, 200, 400, 600, 800, or 1000 nucleotides of nucleotides 1 to 1751 of SEQ ID NO:5. Alternatively, a *ZHMUP-2* gene promoter may be provided by nucleotides 1 to 1751 of SEQ ID NO:5. The identification of genomic fragments containing a *ZHMUP-2* promoter or regulatory element can be achieved using well-established techniques, such as deletion analysis (see, generally, Ausubel (1995)).

Cloning of 5' flanking sequences also facilitates production of *ZHMUP-2* proteins by "gene activation," as disclosed in U.S. Patent No. 5,641,670. Briefly, expression of an endogenous *ZHMUP-2* gene in a cell is altered by introducing into the *ZHMUP-2* locus a DNA construct comprising at least a targeting sequence, a regulatory sequence, an exon, and an unpaired splice donor site. The targeting sequence is a *ZHMUP-2* 5' noncoding sequence that permits homologous recombination of the construct with the endogenous *ZHMUP-2* locus, whereby the sequences within the construct become operably linked with the endogenous *ZHMUP-2* coding sequence. In this way, an endogenous *ZHMUP-2* promoter can be replaced or supplemented with other regulatory sequences to provide enhanced, tissue-specific, or otherwise regulated expression.

#### 4. *Production of ZHMUP-2 Variants*

The present invention provides a variety of nucleic acid molecules, including DNA and RNA molecules, which encode the *ZHMUP-2* polypeptides



disclosed herein. Those skilled in the art will readily recognize that, in view of the degeneracy of the genetic code, considerable sequence variation is possible among these polynucleotide molecules. SEQ ID NO:3 is a degenerate nucleotide sequence that encompasses all nucleic acid molecules that encode the ZHMUP-2 polypeptide of SEQ ID NO:2. Those skilled in the art will recognize that the degenerate sequence of SEQ ID NO:3 also provides all RNA sequences encoding SEQ ID NO:2, by substituting U for T. Thus, the present invention contemplates ZHMUP-2 polypeptide-encoding nucleic acid molecules comprising the nucleotide sequence of SEQ ID NO:1, and their RNA equivalents.

10           Table 2 sets forth the one-letter codes used within SEQ ID NO:3 to denote degenerate nucleotide positions. "Resolutions" are the nucleotides denoted by a code letter. "Complement" indicates the code for the complementary nucleotide(s). For example, the code Y denotes either C or T, and its complement R denotes A or G, A being complementary to T, and G being complementary to C.

Table 2

Nucleotide	Resolution	Complement	Resolution
A	A	T	T
C	C	G	C
G	G	C	C
T	T	A	A
K	A G	Y	C T
Y	C T	R	A G
M	A C	K	G T
K	G T	M	A C
S	C G	S	C G
W	A T	W	A T
H	A C T	D	A G T
B	C G T	V	A C G
V	A C G	B	C G T
D	A G T	H	A C T
N	A C G T	N	A C G T

The degenerate codons used in SEQ ID NO:3, encompassing all possible  
5 codons for a given amino acid, are set forth in Table 3.

Table 3

Amino Acid	One Letter Code	Codons	Degenerate Codon
Cys	C	TGC TGT	TGY
Ser	S	AGC AGT TCA TCC TCG TCT	WSN
Thr	T	ACA ACC ACG ACT	ACN
Pro	P	CCA CCC CCG CCT	CCN
Ala	A	GCA GCC GCG GCT	GCN
Gly	G	GGA GGC GGG GGT	GGN
Asn	N	AAC AAT	AAY
Asp	D	GAC GAT	GAY
Glu	E	GAA GAG	GAR
Gln	Q	CAA CAG	CAR
His	H	CAC CAT	CAY
Arg	R	AGA AGG CGA CGC CGG CGT	MGN
Lys	K	AAA AAG	AAR
Met	M	ATG	ATG
Ile	I	ATA ATC ATT	ATH
Leu	L	CTA CTC CTG CTT TTA TTG	YTN
Val	V	GTA GTC GTG GTT	GTN
Phe	F	TTC TTT	TTY
Tyr	Y	TAC TAT	TAY
Trp	W	TGG	TGG
Ter	.	TAA TAG TGA	TRR
Asn Asp	B		RAY
Glu Gln	Z		SAR
Any	X		NNN

One of ordinary skill in the art will appreciate that some ambiguity is introduced in determining a degenerate codon, representative of all possible codons encoding an amino acid. For example, the degenerate codon for serine (WSN) can, in some circumstances, encode arginine (AGR), and the degenerate codon for arginine (MGN) can, in some circumstances, encode serine (AGY). A similar relationship exists between codons encoding phenylalanine and leucine. Thus, some polynucleotides encompassed by the degenerate sequence may encode variant amino acid sequences, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequence of SEQ ID NO:2. Variant sequences can be readily tested for functionality as described herein.

Different species can exhibit "preferential codon usage." In general, see, Grantham *et al.*, *Nuc. Acids Res.* 8:1893 (1980), Haas *et al.* *Curr. Biol.* 6:315 (1996), Wain-Hobson *et al.*, *Gene* 13:355 (1981), Grosjean and Fiers, *Gene* 18:199 (1982), Holm, *Nuc. Acids Res.* 14:3075 (1986), Ikemura, *J. Mol. Biol.* 158:573 (1982), Sharp and Matassi, *Curr. Opin. Genet. Dev.* 4:851 (1994), Kane, *Curr. Opin. Biotechnol.* 6:494 (1995), and Makrides, *Microbiol. Rev.* 60:512 (1996). As used herein, the term "preferential codon usage" or "preferential codons" is a term of art referring to protein translation codons that are most frequently used in cells of a certain species, thus favoring one or a few representatives of the possible codons encoding each amino acid (see Table 3). For example, the amino acid threonine (Thr) may be encoded by ACA, ACC, ACG, or ACT, but in mammalian cells ACC is the most commonly used codon; in other species, for example, insect cells, yeast, viruses or bacteria, different Thr codons may be preferential. Preferential codons for a particular species can be introduced into the polynucleotides of the present invention by a variety of methods known in the art. Introduction of preferential codon sequences into recombinant DNA can, for example, enhance production of the protein by making protein translation more efficient within a particular cell type or species. Therefore, the degenerate codon sequence disclosed in SEQ ID NO:3 serves as a template for optimizing expression of polynucleotides in various cell types and species commonly used in the art and disclosed herein. Sequences containing preferential codons can be tested and optimized for expression in various species, and tested for functionality as disclosed herein.

The present invention further provides variant polypeptides and nucleic acid molecules that represent counterparts from other species (orthologs). These species include, but are not limited to mammalian, avian, amphibian, reptile, fish, insect and other vertebrate and invertebrate species. Of particular interest are ZHMUP-2 polypeptides from other mammalian species, including porcine, rat, ovine, murine, bovine, canine, feline, equine, and other primate polypeptides. Orthologs of human

ZHMUP-2 can be cloned using information and compositions provided by the present invention in combination with conventional cloning techniques. For example, a cDNA can be cloned using mRNA obtained from a tissue or cell type that expresses ZHMUP-2 as disclosed herein. Suitable sources of mRNA can be identified by probing northern  
5 blots with probes designed from the sequences disclosed herein. A library is then prepared from mRNA of a positive tissue or cell line.

A ZHMUP-2-encoding cDNA can then be isolated by a variety of methods, such as by probing with a complete or partial human cDNA or with one or more sets of degenerate probes based on the disclosed sequences. A cDNA can also be  
10 cloned using the polymerase chain reaction with primers designed from the representative human ZHMUP-2 sequences disclosed herein. Within an additional method, the cDNA library can be used to transform or transfect host cells, and expression of the cDNA of interest can be detected with an antibody to ZHMUP-2 polypeptide. Similar techniques can also be applied to the isolation of genomic clones.

Those skilled in the art will recognize that the sequence disclosed in SEQ  
15 ID NO:1 represents a single allele of human ZHMUP-2, and that allelic variation and alternative splicing are expected to occur. Allelic variants of this sequence can be cloned by probing cDNA or genomic libraries from different individuals according to standard procedures. Allelic variants of the nucleotide sequence shown in SEQ ID NO:1,  
20 including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants of SEQ ID NO:2. cDNA molecules generated from alternatively spliced mRNAs, which retain the properties of the ZHMUP-2 polypeptide are included within the scope of the present invention, as are polypeptides encoded by  
25 such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individuals or tissues according to standard procedures known in the art.

Within certain embodiments of the invention, the isolated nucleic acid molecules can hybridize under stringent conditions to nucleic acid molecules comprising  
30 nucleotide sequences disclosed herein. For example, such nucleic acid molecules can hybridize under stringent conditions to nucleic acid molecules comprising the nucleotide sequence of SEQ ID NO:1, or comprising the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1, or to nucleic acid molecules consisting of a nucleotide sequence that is complementary to such nucleotide sequences. In general, stringent conditions are  
35 selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined

ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe.

A pair of nucleic acid molecules, such as DNA-DNA, RNA-RNA and DNA-RNA, can hybridize if the nucleotide sequences have some degree of complementarity. Hybrids can tolerate mismatched base pairs in the double helix, but the stability of the hybrid is influenced by the degree of mismatch. The  $T_m$  of the mismatched hybrid decreases by 1°C for every 1-1.5% base pair mismatch. Varying the stringency of the hybridization conditions allows control over the degree of mismatch that will be present in the hybrid. The degree of stringency increases as the hybridization temperature increases and the ionic strength of the hybridization buffer decreases. Stringent hybridization conditions encompass temperatures of about 5-25°C below the  $T_m$  of the hybrid and a hybridization buffer having up to 1 M Na<sup>+</sup>. Higher degrees of stringency at lower temperatures can be achieved with the addition of formamide which reduces the  $T_m$  of the hybrid about 1°C for each 1% formamide in the buffer solution. Generally, such stringent conditions include temperatures of 20-70°C and a hybridization buffer containing up to 6x SSC and 0-50% formamide. A higher degree of stringency can be achieved at temperatures of from 40-70°C with a hybridization buffer having up to 4x SSC and from 0-50% formamide. Highly stringent conditions typically encompass temperatures of 42-70°C with a hybridization buffer having up to 1x SSC and 0-50% formamide. Different degrees of stringency can be used during hybridization and washing to achieve maximum specific binding to the target sequence. Typically, the washes following hybridization are performed at increasing degrees of stringency to remove non-hybridized polynucleotide probes from hybridized complexes.

The above conditions are meant to serve as a guide and it is well within the abilities of one skilled in the art to adapt these conditions for use with a particular polypeptide hybrid. The  $T_m$  for a specific target sequence is the temperature (under defined conditions) at which 50% of the target sequence will hybridize to a perfectly matched probe sequence. Those conditions that influence the  $T_m$  include, the size and base pair content of the polynucleotide probe, the ionic strength of the hybridization solution, and the presence of destabilizing agents in the hybridization solution. Numerous equations for calculating  $T_m$  are known in the art, and are specific for DNA, RNA and DNA-RNA hybrids and polynucleotide probe sequences of varying length (see, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition (Cold Spring Harbor Press 1989); Ausubel *et al.*, (eds.), *Current Protocols in Molecular Biology* (John Wiley and Sons, Inc. 1987); Berger and Kimmel (eds.), *Guide to Molecular Cloning Techniques*, (Academic Press, Inc. 1987); and Wetmur, *Crit. Rev. Biochem. Mol. Biol.* 26:227 (1990)). Sequence analysis software such as OLIGO 6.0

(LSR; Long Lake, MN) and *Primer Premier 4.0* (Premier Biosoft International; Palo Alto, CA), as well as sites on the Internet, are available tools for analyzing a given sequence and calculating  $T_m$  based on user defined criteria. Such programs can also analyze a given sequence under defined conditions and identify suitable probe sequences.

- 5 Typically, hybridization of longer polynucleotide sequences, >50 base pairs, is performed at temperatures of about 20-25°C below the calculated  $T_m$ . For smaller probes, <50 base pairs, hybridization is typically carried out at the  $T_m$  or 5-10°C below. This allows for the maximum rate of hybridization for DNA-DNA and DNA-RNA hybrids.

- 10 The length of the polynucleotide sequence influences the rate and stability of hybrid formation. Smaller probe sequences, <50 base pairs, reach equilibrium with complementary sequences rapidly, but may form less stable hybrids. Incubation times of anywhere from minutes to hours can be used to achieve hybrid formation. Longer probe sequences come to equilibrium more slowly, but form more stable complexes even at  
15 lower temperatures. Incubations are allowed to proceed overnight or longer. Generally, incubations are carried out for a period equal to three times the calculated  $Cot$  time.  $Cot$  time, the time it takes for the polynucleotide sequences to reassociate, can be calculated for a particular sequence by methods known in the art.

- The base pair composition of polynucleotide sequence will effect the  
20 thermal stability of the hybrid complex, thereby influencing the choice of hybridization temperature and the ionic strength of the hybridization buffer. A-T pairs are less stable than G-C pairs in aqueous solutions containing sodium chloride. Therefore, the higher the G-C content, the more stable the hybrid. Even distribution of G and C residues within the sequence also contribute positively to hybrid stability. In addition, the base  
25 pair composition can be manipulated to alter the  $T_m$  of a given sequence. For example, 5-methyldeoxycytidine can be substituted for deoxycytidine and 5-bromodeoxuridine can be substituted for thymidine to increase the  $T_m$ , whereas 7-deaz-2'-deoxyguanosine can be substituted for guanosine to reduce dependence on  $T_m$ .

- The ionic concentration of the hybridization buffer also affects the  
30 stability of the hybrid. Hybridization buffers generally contain blocking agents such as Denhardt's solution (Sigma Chemical Co., St. Louis, Mo.), denatured salmon sperm DNA, tRNA, milk powders (BLOTTO), heparin or SDS, and a  $Na^+$  source, such as SSC (1x SSC: 0.15 M sodium chloride, 15 mM sodium citrate) or SSPE (1x SSPE: 1.8 M NaCl, 10 mM  $NaH_2PO_4$ , 1 mM EDTA, pH 7.7). By decreasing the ionic concentration  
35 of the buffer, the stability of the hybrid is increased. Typically, hybridization buffers contain from between 10 mM - 1 M  $Na^+$ . The addition of destabilizing or denaturing agents such as formamide, tetraalkylammonium salts, guanidinium cations or thiocyanate

cations to the hybridization solution will alter the  $T_m$  of a hybrid. Typically, formamide is used at a concentration of up to 50% to allow incubations to be carried out at more convenient and lower temperatures. Formamide also acts to reduce non-specific background when using RNA probes.

5           As an illustration, a nucleic acid molecule encoding a variant ZHMUP-2 polypeptide can be hybridized with a nucleic acid molecule having the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1 (or its complement) at 42°C overnight in a solution comprising 50% formamide, 5xSSC (1xSSC: 0.15 M sodium chloride and 15 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's  
10   solution (100x Denhardt's solution: 2% (w/v) Ficoll 400, 2% (w/v) polyvinylpyrrolidone, and 2% (w/v) bovine serum albumin, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA. One of skill in the art can devise variations of these hybridization conditions. For example, the hybridization mixture can be incubated at a higher temperature, such as about 65°C, in a solution that does not  
15   contain formamide. Moreover, premixed hybridization solutions are available (*e.g.*, EXPRESSHYB Hybridization Solution from CLONTECH Laboratories, Inc.), and hybridization can be performed according to the manufacturer's instructions.

Following hybridization, the nucleic acid molecules can be washed to remove non-hybridized nucleic acid molecules under stringent conditions, or under  
20   highly stringent conditions. Typical stringent washing conditions include washing in a solution of 0.5x - 2x SSC with 0.1% sodium dodecyl sulfate (SDS) at 55 - 65°C. For example, certain nucleic acid molecules encoding a variant ZHMUP-2 polypeptide remain hybridized following stringent washing conditions with a nucleic acid molecule consisting of the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1 (or its  
25   complement), in which the wash stringency is equivalent to 0.5x - 2x SSC with 0.1% SDS at 55 - 65°C, including 0.5x SSC with 0.1% SDS at 55°C, or 2xSSC with 0.1% SDS at 65°C. One of skill in the art can readily devise equivalent conditions, for example, by substituting the SSPE for SSC in the wash solution.

Typical highly stringent washing conditions include washing in a solution  
30   of 0.1x - 0.2x SSC with 0.1% sodium dodecyl sulfate (SDS) at 50 - 65°C. As an illustration, particular nucleic acid molecules encoding a variant ZHMUP-2 polypeptide remain hybridized following stringent washing conditions with a nucleic acid molecule having the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1 (or its complement), in which the wash stringency is equivalent to 0.1x - 0.2x SSC with 0.1%  
35   SDS at 50 - 65°C, including 0.1x SSC with 0.1% SDS at 50°C, or 0.2xSSC with 0.1% SDS at 65°C.



The present invention also provides isolated ZHMUP-2 polypeptides that have a substantially similar sequence identity to the polypeptide of SEQ ID NO:2, or orthologs. The term "substantially similar sequence identity" is used herein to denote polypeptides having 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to the sequence shown in SEQ ID NO:2.

The present invention also contemplates ZHMUP-2 variant nucleic acid molecules that can be identified using two criteria: a determination of the similarity between the encoded polypeptide with the amino acid sequence of SEQ ID NO:2, and a hybridization assay, as described above. Such ZHMUP-2 variants include nucleic acid molecules (1) that remain hybridized following stringent washing conditions with a nucleic acid molecule comprising the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1 (or its complement), in which the wash stringency is equivalent to 0.5x - 2x SSC with 0.1% SDS at 55 - 65°C, and (2) that encode a polypeptide having 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity to the amino acid sequence of SEQ ID NO:2.

Alternatively, ZHMUP-2 variants can be characterized as nucleic acid molecules (1) that remain hybridized following highly stringent washing conditions with a nucleic acid molecule comprising the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1 (or its complement), in which the wash stringency is equivalent to 0.1x - 0.2x SSC with 0.1% SDS at 50 - 65°C, and (2) that encode a polypeptide having 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity to the amino acid sequence of SEQ ID NO:2.

Percent sequence identity is determined by conventional methods. See, for example, Altschul *et al.*, *Bull. Math. Bio.* 48:603 (1986), and Henikoff and Henikoff, *Proc. Nat'l Acad. Sci. USA* 89:10915 (1992). Briefly, two amino acid sequences are aligned to optimize the alignment scores using a gap opening penalty of 10, a gap extension penalty of 1, and the "BLOSUM62" scoring matrix of Henikoff and Henikoff (*ibid.*) as shown in Table 4 (amino acids are indicated by the standard one-letter codes). The percent identity is then calculated as:  $([\text{Total number of identical matches}] / [\text{length of the longer sequence plus the number of gaps introduced into the longer sequence in order to align the two sequences}]) (100)$ .

Table 4

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

5

10

15

20

Those skilled in the art appreciate that there are many established algorithms available to align two amino acid sequences. The "FASTA" similarity search algorithm of Pearson and Lipman is a suitable protein alignment method for examining the level of identity shared by an amino acid sequence disclosed herein and the amino acid sequence of a putative ZHMUP-2 variant. The FASTA algorithm is described by Pearson and Lipman, *Proc. Nat'l Acad. Sci. USA* 85:2444 (1988), and by Pearson, *Meth. Enzymol.* 183:63 (1990). Briefly, FASTA first characterizes sequence similarity by identifying regions shared by the query sequence (*e.g.*, SEQ ID NO:2) and a test sequence that have either the highest density of identities (if the ktup variable is 1) or pairs of identities (if ktup=2), without considering conservative amino acid substitutions, insertions, or deletions. The ten regions with the highest density of identities are then rescored by comparing the similarity of all paired amino acids using an amino acid substitution matrix, and the ends of the regions are "trimmed" to include only those residues that contribute to the highest score. If there are several regions with scores greater than the "cutoff" value (calculated by a predetermined formula based upon the length of the sequence and the ktup value), then the trimmed initial regions are examined to determine whether the regions can be joined to form an approximate alignment with gaps. Finally, the highest scoring regions of the two amino acid sequences are aligned using a modification of the Needleman-Wunsch-Sellers algorithm (Needleman and Wunsch, *J. Mol. Biol.* 48:444 (1970); Sellers, *SIAM J. Appl. Math.* 26:787 (1974)), which allows for amino acid insertions and deletions. Illustrative parameters for FASTA analysis are: ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=BLOSUM62. These parameters can be introduced into a FASTA program by modifying the scoring matrix file ("SMATRIX"), as explained in Appendix 2 of Pearson, *Meth. Enzymol.* 183:63 (1990).

FASTA can also be used to determine the sequence identity of nucleic acid molecules using a ratio as disclosed above. For nucleotide sequence comparisons, the ktup value can range between one to six, preferably from three to six, most preferably three, with other parameters set as described above.

The present invention includes nucleic acid molecules that encode a polypeptide having a conservative amino acid change, compared with the amino acid sequence of SEQ ID NO:2. That is, variants can be obtained that contain one or more amino acid substitutions of SEQ ID NO:2, in which an alkyl amino acid is substituted for an alkyl amino acid in a ZHMUP-2 amino acid sequence, an aromatic amino acid is substituted for an aromatic amino acid in a ZHMUP-2 amino acid sequence, a sulfur-containing amino acid is substituted for a sulfur-containing amino acid in a ZHMUP-2 amino acid sequence, a hydroxy-containing amino acid is substituted for a hydroxy-

containing amino acid in a ZHMUP-2 amino acid sequence, an acidic amino acid is substituted for an acidic amino acid in a ZHMUP-2 amino acid sequence, a basic amino acid is substituted for a basic amino acid in a ZHMUP-2 amino acid sequence, or a dibasic monocarboxylic amino acid is substituted for a dibasic monocarboxylic amino acid in a ZHMUP-2 amino acid sequence.

Among the common amino acids, for example, a "conservative amino acid substitution" is illustrated by a substitution among amino acids within each of the following groups: (1) glycine, alanine, valine, leucine, and isoleucine, (2) phenylalanine, tyrosine, and tryptophan, (3) serine and threonine, (4) aspartate and glutamate, (5) glutamine and asparagine, and (6) lysine, arginine and histidine.

The BLOSUM62 table is an amino acid substitution matrix derived from about 2,000 local multiple alignments of protein sequence segments, representing highly conserved regions of more than 500 groups of related proteins (Henikoff and Henikoff, *Proc. Nat'l Acad. Sci. USA* 89:10915 (1992)). Accordingly, the BLOSUM62 substitution frequencies can be used to define conservative amino acid substitutions that may be introduced into the amino acid sequences of the present invention. Although it is possible to design amino acid substitutions based solely upon chemical properties (as discussed above), the language "conservative amino acid substitution" preferably refers to a substitution represented by a BLOSUM62 value of greater than -1. For example, an amino acid substitution is conservative if the substitution is characterized by a BLOSUM62 value of 0, 1, 2, or 3. According to this system, preferred conservative amino acid substitutions are characterized by a BLOSUM62 value of at least 1 (e.g., 1, 2 or 3), while more preferred conservative amino acid substitutions are characterized by a BLOSUM62 value of at least 2 (e.g., 2 or 3).

Particular variants of ZHMUP-2 are characterized by having greater than 96%, at least 97%, at least 98%, or at least 99% sequence identity to the corresponding amino acid sequence (e.g., amino acid residues 16 to 185 of SEQ ID NO:2), wherein the variation in amino acid sequence is due to one or more conservative amino acid substitutions.

Conservative amino acid changes in a ZHMUP-2 gene can be introduced by substituting nucleotides for the nucleotides recited in SEQ ID NO:1. Such "conservative amino acid" variants can be obtained, for example, by oligonucleotide-directed mutagenesis, linker-scanning mutagenesis, mutagenesis using the polymerase chain reaction, and the like (see Ausubel (1995) at pages 8-10 to 8-22; and McPherson (ed.), *Directed Mutagenesis: A Practical Approach* (IRL Press 1991)).

The proteins of the present invention can also comprise non-naturally occurring amino acid residues. Non-naturally occurring amino acids include, without

limitation, *trans*-3-methylproline, 2,4-methanoproline, *cis*-4-hydroxyproline, *trans*-4-hydroxyproline, *N*-methylglycine, *allo*-threonine, methylthreonine, hydroxyethylcysteine, hydroxyethylhomocysteine, nitroglutamine, homoglutamine, pipecolic acid, thiazolidine carboxylic acid, dehydroproline, 3- and 4-methylproline, 3,3-dimethylproline, *tert*-leucine, norvaline, 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, and 4-fluorophenylalanine. Several methods are known in the art for incorporating non-naturally occurring amino acid residues into proteins. For example, an *in vitro* system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plasmids containing nonsense mutations is typically carried out in a cell-free system comprising an *E. coli* S30 extract and commercially available enzymes and other reagents. Proteins are purified by chromatography. See, for example, Robertson *et al.*, *J. Am. Chem. Soc.* 113:2722 (1991), Ellman *et al.*, *Methods Enzymol.* 202:301 (1991), Chung *et al.*, *Science* 259:806 (1993), and Chung *et al.*, *Proc. Nat'l Acad. Sci. USA* 90:10145 (1993).

In a second method, translation is carried out in *Xenopus* oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti *et al.*, *J. Biol. Chem.* 271:19991 (1996)). Within a third method, *E. coli* cells are cultured in the absence of a natural amino acid that is to be replaced (*e.g.*, phenylalanine) and in the presence of the desired non-naturally occurring amino acid(s) (*e.g.*, 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occurring amino acid is incorporated into the protein in place of its natural counterpart. See, Koide *et al.*, *Biochem.* 33:7470 (1994). Naturally occurring amino acid residues can be converted to non-naturally occurring species by *in vitro* chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, *Protein Sci.* 2:395 (1993)).

A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, non-naturally occurring amino acids, and unnatural amino acids may be substituted for ZHMUP-2 amino acid residues.

Essential amino acids in the polypeptides of the present invention can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science* 244:1081 (1989), Bass *et al.*, *Proc. Nat'l Acad. Sci. USA* 88:4498 (1991), Coombs and Corey, "Site-Directed Mutagenesis and Protein Engineering," in *Proteins: Analysis and Design*, Angeletti (ed.), pages 259-311 (Academic Press, Inc. 1998)). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant mutant

molecules are tested for biological activity as disclosed below to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton *et al.*, *J. Biol. Chem.* 271:4699 (1996).

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening, such as those disclosed by Reidhaar-Olson and Sauer (*Science* 241:53 (1988)) or Bowie and Sauer (*Proc. Nat'l Acad. Sci. USA* 86:2152 (1989)). Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (*e.g.*, Lowman *et al.*, *Biochem.* 30:10832 (1991), Ladner *et al.*, U.S. Patent No. 5,223,409, Huse, international publication No. WO 92/06204, and region-directed mutagenesis (Derbyshire *et al.*, *Gene* 46:145 (1986), and Ner *et al.*, *DNA* 7:127, (1988)).

Variants of the disclosed ZHMUP-2 nucleotide and polypeptide sequences can also be generated through DNA shuffling as disclosed by Stemmer, *Nature* 370:389 (1994), Stemmer, *Proc. Nat'l Acad. Sci. USA* 91:10747 (1994), and international publication No. WO 97/20078. Briefly, variant DNAs are generated by *in vitro* homologous recombination by random fragmentation of a parent DNA followed by reassembly using PCR, resulting in randomly introduced point mutations. This technique can be modified by using a family of parent DNAs, such as allelic variants or DNAs from different species, to introduce additional variability into the process. Selection or screening for the desired activity, followed by additional iterations of mutagenesis and assay provides for rapid "evolution" of sequences by selecting for desirable mutations while simultaneously selecting against detrimental changes.

Mutagenesis methods as disclosed herein can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides in host cells. Mutagenized DNA molecules that encode biologically active polypeptides, or polypeptides that bind with anti-ZHMUP-2 antibodies, can be recovered from the host cells and rapidly sequenced using modern equipment. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide of interest, and can be applied to polypeptides of unknown structure.

The present invention also includes "functional fragments" of ZHMUP-2 polypeptides and nucleic acid molecules encoding such functional fragments. Routine deletion analyses of nucleic acid molecules can be performed to obtain functional fragments of a nucleic acid molecule that encodes a ZHMUP-2 polypeptide. As an illustration, DNA molecules having the nucleotide sequence of SEQ ID NO:1 can be digested with *Bal31* nuclease to obtain a series of nested deletions. One alternative to

exonuclease digestion is to use oligonucleotide-directed mutagenesis to introduce deletions or stop codons to specify production of a desired fragment. Alternatively, particular fragments of a *ZHMUP-2* gene can be synthesized using the polymerase chain reaction.

5           As an illustration, studies on the truncation at either or both termini of interferons have been summarized by Horisberger and Di Marco, *Pharmac. Ther.* 66:507 (1995). Moreover, standard techniques for functional analysis of proteins are described by, for example, Treuter *et al.*, *Molec. Gen. Genet.* 240:113 (1993), Content *et al.*, "Expression and preliminary deletion analysis of the 42 kDa 2-5A synthetase induced by human interferon," in *Biological Interferon Systems, Proceedings of ISIR-TNO Meeting on Interferon Systems*, Cantell (ed.), pages 65-72 (Nijhoff 1987), Herschman, "The EGF Receptor," in *Control of Animal Cell Proliferation, Vol. 1*, Boynton *et al.*, (eds.) pages 169-199 (Academic Press 1985), Coumailleau *et al.*, *J. Biol. Chem.* 270:29270 (1995); Fukunaga *et al.*, *J. Biol. Chem.* 270:25291 (1995); Yamaguchi *et al.*, *Biochem. Pharmacol.* 50:1295 (1995), and Meisel *et al.*, *Plant Molec. Biol.* 30:1 (1996).

15           The present invention also contemplates functional fragments of a *ZHMUP-2* gene that has amino acid changes, compared with the amino acid sequence of SEQ ID NO:2. A variant *ZHMUP-2* gene can be identified on the basis of structure by determining the level of identity with nucleotide and amino acid sequences of SEQ ID NOs:1 and 2, as discussed above. An alternative approach to identifying a variant gene on the basis of structure is to determine whether a nucleic acid molecule encoding a potential variant *ZHMUP-2* gene can hybridize to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, as discussed above.

20           The present invention also provides polypeptide fragments or peptides comprising an epitope-bearing portion of a *ZHMUP-2* polypeptide described herein. Such fragments or peptides may comprise an "immunogenic epitope," which is a part of a protein that elicits an antibody response when the entire protein is used as an immunogen. Immunogenic epitope-bearing peptides can be identified using standard methods (see, for example, Geysen *et al.*, *Proc. Nat'l Acad. Sci. USA* 81:3998 (1983)).

25           In contrast, polypeptide fragments or peptides may comprise an "antigenic epitope," which is a region of a protein molecule to which an antibody can specifically bind. Certain epitopes consist of a linear or contiguous stretch of amino acids, and the antigenicity of such an epitope is not disrupted by denaturing agents. It is known in the art that relatively short synthetic peptides that can mimic epitopes of a protein can be used to stimulate the production of antibodies against the protein (see, for example, Sutcliffe *et al.*, *Science* 219:660 (1983)). Accordingly, antigenic epitope-bearing

peptides and polypeptides of the present invention are useful to raise antibodies that bind with the polypeptides described herein.

Antigenic epitope-bearing peptides and polypeptides can contain at least four to ten amino acids, at least ten to fifteen amino acids, or about 15 to about 30 amino acids of SEQ ID NO:2. Such epitope-bearing peptides and polypeptides can be produced by fragmenting a ZHMUP-2 polypeptide, or by chemical peptide synthesis, as described herein. Moreover, epitopes can be selected by phage display of random peptide libraries (see, for example, Lane and Stephen, *Curr. Opin. Immunol.* 5:268 (1993), and Cortese *et al.*, *Curr. Opin. Biotechnol.* 7:616 (1996)). Standard methods for identifying epitopes and producing antibodies from small peptides that comprise an epitope are described, for example, by Mole, "Epitope Mapping," in *Methods in Molecular Biology*, Vol. 10, Manson (ed.), pages 105-116 (The Humana Press, Inc. 1992), Price, "Production and Characterization of Synthetic Peptide-Derived Antibodies," in *Monoclonal Antibodies: Production, Engineering, and Clinical Application*, Ritter and Ladyman (eds.), pages 60-84 (Cambridge University Press 1995), and Coligan *et al.* (eds.), *Current Protocols in Immunology*, pages 9.3.1 - 9.3.5 and pages 9.4.1 - 9.4.11 (John Wiley & Sons 1997).

For any ZHMUP-2 polypeptide, including variants and fusion proteins, one of ordinary skill in the art can readily generate a fully degenerate polynucleotide sequence encoding that variant using the information set forth in Tables 2 and 3 above. Moreover, those of skill in the art can use standard software to devise ZHMUP-2 variants based upon the nucleotide and amino acid sequences described herein. Accordingly, the present invention includes a computer-readable medium encoded with a data structure that provides at least one of SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3. Suitable forms of computer-readable media include magnetic media and optically-readable media. Examples of magnetic media include a hard or fixed drive, a random access memory (RAM) chip, a floppy disk, digital linear tape (DLT), a disk cache, and a ZIP disk. Optically readable media are exemplified by compact discs (*e.g.*, CD-read only memory (ROM), CD-rewritable (RW), and CD-recordable), and digital versatile/video discs (DVD) (*e.g.*, DVD-ROM, DVD-RAM, and DVD+RW).

## 5. Production of ZHMUP-2 Fusion Proteins

Fusion proteins of ZHMUP-2 can be used to express ZHMUP-2 in a recombinant host, and to isolate expressed ZHMUP-2. One type of fusion protein comprises a peptide that guides a ZHMUP-2 polypeptide from a recombinant host cell. To direct a ZHMUP-2 polypeptide into the secretory pathway of a eukaryotic host cell, a secretory signal sequence (also known as a signal peptide, a leader sequence, prepro



sequence or pre sequence) is provided in the ZHMUP-2 expression vector. While the secretory signal sequence may be derived from ZHMUP-2, a suitable signal sequence may also be derived from another secreted protein or synthesized *de novo*. The secretory signal sequence is operably linked to a ZHMUP-2-encoding sequence such that the two sequences are joined in the correct reading frame and positioned to direct the newly synthesized polypeptide into the secretory pathway of the host cell. Secretory signal sequences are commonly positioned 5' to the nucleotide sequence encoding the polypeptide of interest, although certain secretory signal sequences may be positioned elsewhere in the nucleotide sequence of interest (see, *e.g.*, Welch *et al.*, U.S. Patent No. 5,037,743; Holland *et al.*, U.S. Patent No. 5,143,830).

While the secretory signal sequence of ZHMUP-2 or another protein produced by mammalian cells (*e.g.*, tissue-type plasminogen activator signal sequence, as described, for example, in U.S. Patent No. 5,641,655) is useful for expression of ZHMUP-2 in recombinant mammalian hosts, a yeast signal sequence is preferred for expression in yeast cells. Examples of suitable yeast signal sequences are those derived from yeast mating pheromone  $\alpha$ -factor (encoded by the *MF $\alpha$ 1* gene), invertase (encoded by the *SUC2* gene), or acid phosphatase (encoded by the *PHO5* gene). See, for example, Romanos *et al.*, "Expression of Cloned Genes in Yeast," in *DNA Cloning 2: A Practical Approach*, 2<sup>nd</sup> Edition, Glover and Hames (eds.), pages 123-167 (Oxford University Press 1995).

In bacterial cells, it is often desirable to express a heterologous protein as a fusion protein to decrease toxicity, increase stability, and to enhance recovery of the expressed protein. For example, ZHMUP-2 can be expressed as a fusion protein comprising a glutathione S-transferase polypeptide. Glutathione S-transferase fusion proteins are typically soluble, and easily purifiable from *E. coli* lysates on immobilized glutathione columns. In similar approaches, a ZHMUP-2 fusion protein comprising a maltose binding protein polypeptide can be isolated with an amylose resin column, while a fusion protein comprising the C-terminal end of a truncated Protein A gene can be purified using IgG-Sepharose. Established techniques for expressing a heterologous polypeptide as a fusion protein in a bacterial cell are described, for example, by Williams *et al.*, "Expression of Foreign Proteins in *E. coli* Using Plasmid Vectors and Purification of Specific Polyclonal Antibodies," in *DNA Cloning 2: A Practical Approach*, 2<sup>nd</sup> Edition, Glover and Hames (Eds.), pages 15-58 (Oxford University Press 1995). In addition, commercially available expression systems are available. For example, the PINPOINT Xa protein purification system (Promega Corporation; Madison, WI) provides a method for isolating a fusion protein comprising a polypeptide that becomes biotinylated during expression with a resin that comprises avidin.

Peptide tags that are useful for isolating heterologous polypeptides expressed by either prokaryotic or eukaryotic cells include polyHistidine tags (which have an affinity for nickel-chelating resin), *c-myc* tags, calmodulin binding protein (isolated with calmodulin affinity chromatography), substance P, the RYIRS tag (which binds with anti-RYIRS antibodies), the Glu-Glu tag, and the FLAG tag (which binds with anti-FLAG antibodies). See, for example, Luo *et al.*, *Arch. Biochem. Biophys.* 329:215 (1996), Morganti *et al.*, *Biotechnol. Appl. Biochem.* 23:67 (1996), and Zheng *et al.*, *Gene* 186:55 (1997). Nucleic acid molecules encoding such peptide tags are available, for example, from Sigma-Aldrich Corporation (St. Louis, MO).

Another form of fusion protein comprises a ZHMUP-2 polypeptide and an immunoglobulin heavy chain constant region, typically an F<sub>C</sub> fragment, which contains two constant region domains and a hinge region but lacks the variable region. As an illustration, Chang *et al.*, U.S. Patent No. 5,723,125, describe a fusion protein comprising a human interferon and a human immunoglobulin Fc fragment, in which the C-terminal of the interferon is linked to the N-terminal of the Fc fragment by a peptide linker moiety. An example of a peptide linker is a peptide comprising primarily a T cell inert sequence, which is immunologically inert. An exemplary peptide linker has the amino acid sequence: GSGGG SGGGG S (SEQ ID NO:4). In such a fusion protein, an illustrative Fc moiety is a human  $\gamma$ 4 chain, which is stable in solution and has little or no complement activating activity. Accordingly, the present invention contemplates a ZHMUP-2 fusion protein that comprises a ZHMUP-2 moiety and a human Fc fragment, wherein the C-terminus of the ZHMUP-2 moiety is attached to the N-terminus of the Fc fragment via a peptide linker, such as a peptide consisting of the amino acid sequence of SEQ ID NO:4. The ZHMUP-2 moiety can be a ZHMUP-2 molecule or a fragment thereof.

In another variation, a ZHMUP-2 fusion protein comprises an IgG sequence, a ZHMUP-2 moiety covalently joined to the aminoterminal end of the IgG sequence, and a signal peptide that is covalently joined to the aminoterminal of the ZHMUP-2 moiety, wherein the IgG sequence consists of the following elements in the following order: a hinge region, a CH<sub>2</sub> domain, and a CH<sub>3</sub> domain. Accordingly, the IgG sequence lacks a CH<sub>1</sub> domain. The ZHMUP-2 moiety displays a ZHMUP-2 activity, as described herein, such as the ability to bind with a ZHMUP-2 antibody. This general approach to producing fusion proteins that comprise both antibody and nonantibody portions has been described by LaRoche *et al.*, EP 742830 (WO 95/21258).

Fusion proteins comprising a ZHMUP-2 moiety and an Fc moiety can be used, for example, as an *in vitro* assay tool. For example, the presence of a ZHMUP-2 receptor in a biological sample can be detected using a ZHMUP-2-antibody fusion

protein, in which the ZHMUP-2 moiety is used to target the cognate receptor, and a macromolecule, such as Protein A or anti-Fc antibody, is used to detect the bound fusion protein-receptor complex. Furthermore, such fusion proteins can be used to identify agonists and antagonists that interfere with the binding of ZHMUP-2 to its receptor.

5           The present invention also contemplates the use of the secretory signal sequence contained in the ZHMUP-2 polypeptides of the present invention to direct other polypeptides into the secretory pathway. A signal fusion polypeptide can be made wherein a secretory signal sequence, comprising amino acid residues 1 to about 15 of SEQ ID NO:2, is operably linked to another polypeptide using methods known in the art and disclosed herein.

10           Such constructs comprising a ZHMUP-2 secretory signal sequence have numerous applications known in the art. For example, these novel ZHMUP-2 secretory signal sequence fusion constructs can direct the secretion of an active component of a normally non-secreted protein, such as a receptor. Fusion proteins comprising a  
15           ZHMUP-2 signal sequence may be used in a transgenic animal or in a cultured recombinant host to direct polypeptides through the secretory pathway. With regard to the latter, exemplary polypeptides include pharmaceutically active molecules such as Factor VIIa, proinsulin, insulin, follicle stimulating hormone, tissue type plasminogen activator, tumor necrosis factor, interleukins (*e.g.*, interleukin-1 (IL-1), IL-2, IL-3, IL-4,  
20           IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, and IL-19), colony stimulating factors (*e.g.*, granulocyte-colony stimulating factor (G-CSF) and granulocyte macrophage-colony stimulating factor (GM-CSF)), interferons (*e.g.*, interferons- $\alpha$ , - $\beta$ , - $\gamma$ , - $\omega$ ,  $\delta$ , and - $\tau$ ), the stem cell growth factor designated "S1 factor," erythropoietin, and thrombopoietin. The ZHMUP-2 secretory signal sequence  
25           contained in the fusion polypeptides of the present invention is preferably fused amino-terminally to an additional peptide to direct the additional peptide into the secretory pathway.

30           Fusion proteins can be prepared by methods known to those skilled in the art by preparing each component of the fusion protein and chemically conjugating the components. Alternatively, a polynucleotide encoding both components of the fusion protein in the proper reading frame can be generated using known techniques and expressed by the methods described herein. General methods for enzymatic and chemical cleavage of fusion proteins are described, for example, by Ausubel (1995) at pages 16-19 to 16-25.

## 6. *Production of ZHMUP-2 Polypeptides in Cultured Cells*

The polypeptides of the present invention, including full-length polypeptides, functional fragments, and fusion proteins, can be produced in recombinant host cells following conventional techniques. To express a nucleotide sequence that encodes ZHMUP-2, a nucleic acid molecule encoding the polypeptide must be operably linked to regulatory sequences that control transcriptional expression in an expression vector and then, introduced into a host cell. In addition to transcriptional regulatory sequences, such as promoters and enhancers, expression vectors can include translational regulatory sequences and a marker gene, which is suitable for selection of cells that carry the expression vector.

Expression vectors that are suitable for production of a foreign protein in eukaryotic cells typically contain (1) prokaryotic DNA elements coding for a bacterial replication origin and an antibiotic resistance marker to provide for the growth and selection of the expression vector in a bacterial host; (2) eukaryotic DNA elements that control initiation of transcription, such as a promoter; and (3) DNA elements that control the processing of transcripts, such as a transcription termination/polyadenylation sequence. As discussed above, expression vectors can also include nucleotide sequences encoding a secretory sequence that directs the heterologous polypeptide into the secretory pathway of a host cell. For example, a ZHMUP-2 expression vector may comprise a ZHMUP-2 gene and a secretory sequence derived from a ZHMUP-2 gene or another secreted gene.

Expression of ZHMUP-2 can be achieved using nucleic acid molecules that either include or do not include noncoding portions of the ZHMUP-2 gene. However, higher efficiency of production from certain recombinant host cells may be obtained when at least one ZHMUP-2 intron sequence is included in the expression vector.

ZHMUP-2 proteins of the present invention may be expressed in mammalian cells. Examples of suitable mammalian host cells include African green monkey kidney cells (Vero; ATCC CRL 1587), human embryonic kidney cells (293-HEK; ATCC CRL 1573), baby hamster kidney cells (BHK-21, BHK-570; ATCC CRL 8544, ATCC CRL 10314), canine kidney cells (MDCK; ATCC CCL 34), Chinese hamster ovary cells (CHO-K1; ATCC CCL61; CHO DG44 (Chasin *et al.*, *Som. Cell. Molec. Genet.* 12:555, 1986)), rat pituitary cells (GH1; ATCC CCL82), HeLa S3 cells (ATCC CCL2.2), rat hepatoma cells (H-4-II-E; ATCC CRL 1548) SV40-transformed monkey kidney cells (COS-1; ATCC CRL 1650) and murine embryonic cells (NIH-3T3; ATCC CRL 1658).

For a mammalian host, the transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, simian virus, or the like, in which the regulatory signals are associated with a particular gene which has a high level of expression. Suitable transcriptional and translational regulatory sequences also can be obtained from mammalian genes, such as actin, collagen, myosin, and metallothionein genes.

Transcriptional regulatory sequences include a promoter region sufficient to direct the initiation of RNA synthesis. Suitable eukaryotic promoters include the promoter of the mouse *metallothionein I* gene (Hamer *et al.*, *J. Molec. Appl. Genet.* 1:273 (1982)), the *TK* promoter of *Herpes* virus (McKnight, *Cell* 31:355 (1982)), the *SV40* early promoter (Benoist *et al.*, *Nature* 290:304 (1981)), the *Rous* sarcoma virus promoter (Gorman *et al.*, *Proc. Nat'l Acad. Sci. USA* 79:6777 (1982)), the cytomegalovirus promoter (Foecking *et al.*, *Gene* 45:101 (1980)), and the mouse mammary tumor virus promoter (see, generally, Etcheverry, "Expression of Engineered Proteins in Mammalian Cell Culture," in *Protein Engineering: Principles and Practice*, Cleland *et al.* (eds.), pages 163-181 (John Wiley & Sons, Inc. 1996)).

Alternatively, a prokaryotic promoter, such as the bacteriophage T3 RNA polymerase promoter, can be used to control *ZHMUP-2* gene expression in mammalian cells if the prokaryotic promoter is regulated by a eukaryotic promoter (Zhou *et al.*, *Mol. Cell. Biol.* 10:4529 (1990), and Kaufman *et al.*, *Nucl. Acids Res.* 19:4485 (1991)).

An expression vector can be introduced into host cells using a variety of standard techniques including calcium phosphate transfection, liposome-mediated transfection, microprojectile-mediated delivery, electroporation, and the like. The transfected cells can be selected and propagated to provide recombinant host cells that comprise the expression vector stably integrated in the host cell genome. Techniques for introducing vectors into eukaryotic cells and techniques for selecting such stable transformants using a dominant selectable marker are described, for example, by Ausubel (1995) and by Murray (ed.), *Gene Transfer and Expression Protocols* (Humana Press 1991).

For example, one suitable selectable marker is a gene that provides resistance to the antibiotic neomycin. In this case, selection is carried out in the presence of a neomycin-type drug, such as G-418 or the like. Selection systems can also be used to increase the expression level of the gene of interest, a process referred to as "amplification." Amplification is carried out by culturing transfectants in the presence of a low level of the selective agent and then increasing the amount of selective agent to select for cells that produce high levels of the products of the introduced genes. An exemplary amplifiable selectable marker is dihydrofolate reductase, which confers

resistance to methotrexate. Other drug resistance genes (*e.g.*, hygromycin resistance, multi-drug resistance, puromycin acetyltransferase) can also be used. Alternatively, markers that introduce an altered phenotype, such as green fluorescent protein, or cell surface proteins (*e.g.*, CD4, CD8, Class I MHC, and placental alkaline phosphatase) may be used to sort transfected cells from untransfected cells by such means as FACS sorting or magnetic bead separation technology.

ZHMUP-2 polypeptides can also be produced by cultured cells using a viral delivery system. Exemplary viruses for this purpose include adenovirus, herpesvirus, vaccinia virus and adeno-associated virus (AAV). Adenovirus, a double-stranded DNA virus, is currently the best studied gene transfer vector for delivery of heterologous nucleic acid (for a review, see Becker *et al.*, *Meth. Cell Biol.* 43:161 (1994), and Douglas and Curiel, *Science & Medicine* 4:44 (1997)). Advantages of the adenovirus system include the accommodation of relatively large DNA inserts, the ability to grow to high-titer, the ability to infect a broad range of mammalian cell types, and flexibility that allows use with a large number of available vectors containing different promoters.

By deleting portions of the adenovirus genome, larger inserts (up to 7 kb) of heterologous DNA can be accommodated. These inserts can be incorporated into the viral DNA by direct ligation or by homologous recombination with a co-transfected plasmid. An option is to delete the essential *E1* gene from the viral vector, which results in the inability to replicate unless the *E1* gene is provided by the host cell. For example, adenovirus vector infected human 293 cells (ATCC Nos. CRL-1573, 45504, 45505) can be grown as adherent cells or in suspension culture at relatively high cell density to produce significant amounts of protein (see Garnier *et al.*, *Cytotechnol.* 15:145 (1994)).

ZHMUP-2 genes may also be expressed in other higher eukaryotic cells, such as avian, fungal, insect, yeast, or plant cells. The baculovirus system provides an efficient means to introduce cloned ZHMUP-2 genes into insect cells. Suitable expression vectors are based upon the *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV), and contain well-known promoters such as *Drosophila* heat shock protein (hsp) 70 promoter, *Autographa californica* nuclear polyhedrosis virus immediate-early gene promoter (*ie-1*) and the delayed early 39K promoter, baculovirus p10 promoter, and the *Drosophila* metallothionein promoter. A second method of making recombinant baculovirus utilizes a transposon-based system described by Luckow (Luckow, *et al.*, *J. Virol.* 67:4566 (1993)). This system, which utilizes transfer vectors, is sold in the BAC-to-BAC kit (Life Technologies, Rockville, MD). This system utilizes a transfer vector, PFASTBAC (Life Technologies) containing a Tn7 transposon to move the DNA encoding the ZHMUP-2 polypeptide into a baculovirus genome

maintained in *E. coli* as a large plasmid called a "bacmid." See, Hill-Perkins and Possee, *J. Gen. Virol.* 71:971 (1990), Bonning, *et al.*, *J. Gen. Virol.* 75:1551 (1994), and Chazenbalk, and Rapoport, *J. Biol. Chem.* 270:1543 (1995). In addition, transfer vectors can include an in-frame fusion with DNA encoding an epitope tag at the C- or N-terminus of the expressed ZHMUP-2 polypeptide, for example, a Glu-Glu epitope tag (Grussenmeyer *et al.*, *Proc. Nat'l Acad. Sci.* 82:7952 (1985)). Using a technique known in the art, a transfer vector containing a ZHMUP-2 gene is transformed into *E. coli*, and screened for bacmids which contain an interrupted *lacZ* gene indicative of recombinant baculovirus. The bacmid DNA containing the recombinant baculovirus genome is then isolated using common techniques.

The illustrative PFASTBAC vector can be modified to a considerable degree. For example, the polyhedrin promoter can be removed and substituted with the baculovirus basic protein promoter (also known as P<sub>cor</sub>, p6.9 or MP promoter) which is expressed earlier in the baculovirus infection, and has been shown to be advantageous for expressing secreted proteins (see, for example, Hill-Perkins and Possee, *J. Gen. Virol.* 71:971 (1990), Bonning, *et al.*, *J. Gen. Virol.* 75:1551 (1994), and Chazenbalk and Rapoport, *J. Biol. Chem.* 270:1543 (1995). In such transfer vector constructs, a short or long version of the basic protein promoter can be used. Moreover, transfer vectors can be constructed which replace the native ZHMUP-2 secretory signal sequences with secretory signal sequences derived from insect proteins. For example, a secretory signal sequence from Ecdysteroid Glucosyltransferase (EGT), honey bee Melittin (Invitrogen Corporation; Carlsbad, CA), or baculovirus gp67 (PharMingen; San Diego, CA) can be used in constructs to replace the native ZHMUP-2 secretory signal sequence.

The recombinant virus or bacmid is used to transfect host cells. Suitable insect host cells include cell lines derived from IPLB-Sf-21, a *Spodoptera frugiperda* pupal ovarian cell line, such as Sf9 (ATCC CRL 1711), Sf21AE, and Sf21 (Invitrogen Corporation; San Diego, CA), as well as *Drosophila* Schneider-2 cells, and the HIGH FIVEO cell line (Invitrogen) derived from *Trichoplusia ni* (U.S. Patent No. 5,300,435). Commercially available serum-free media can be used to grow and to maintain the cells. Suitable media are Sf900 II™ (Life Technologies) or ESF 921™ (Expression Systems) for the Sf9 cells; and Ex-cello405™ (JRH Biosciences, Lenexa, KS) or Express FiveO™ (Life Technologies) for the *T. ni* cells. When recombinant virus is used, the cells are typically grown up from an inoculation density of approximately 2-5 x 10<sup>5</sup> cells to a density of 1-2 x 10<sup>6</sup> cells at which time a recombinant viral stock is added at a multiplicity of infection (MOD) of 0.1 to 10, more typically near 3.

Established techniques for producing recombinant proteins in baculovirus systems are provided by Bailey *et al.*, "Manipulation of Baculovirus Vectors," in

*Methods in Molecular Biology, Volume 7: Gene Transfer and Expression Protocols*, Murray (ed.), pages 147-168 (The Humana Press, Inc. 1991), by Patel *et al.*, "The baculovirus expression system," in *DNA Cloning 2: Expression Systems, 2nd Edition*, Glover *et al.* (eds.), pages 205-244 (Oxford University Press 1995), by Ausubel (1995) at pages 16-37 to 16-57, by Richardson (ed.), *Baculovirus Expression Protocols* (The Humana Press, Inc. 1995), and by Lucknow, "Insect Cell Expression Technology," in *Protein Engineering: Principles and Practice*, Cleland *et al.* (eds.), pages 183-218 (John Wiley & Sons, Inc. 1996).

Fungal cells, including yeast cells, can also be used to express the genes described herein. Yeast species of particular interest in this regard include *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Pichia methanolica*. Suitable promoters for expression in yeast include promoters from *GALI* (galactose), *PGK* (phosphoglycerate kinase), *ADH* (alcohol dehydrogenase), *AOXI* (alcohol oxidase), *HIS4* (histidinol dehydrogenase), and the like. Many yeast cloning vectors have been designed and are readily available. These vectors include YIp-based vectors, such as YIp5, YRp vectors, such as YRp17, YE<sub>p</sub> vectors such as YE<sub>p</sub>13 and YC<sub>p</sub> vectors, such as YC<sub>p</sub>19. Methods for transforming *S. cerevisiae* cells with exogenous DNA and producing recombinant polypeptides therefrom are disclosed by, for example, Kawasaki, U.S. Patent No. 4,599,311, Kawasaki *et al.*, U.S. Patent No. 4,931,373, Brake, U.S. Patent No. 4,870,008, Welch *et al.*, U.S. Patent No. 5,037,743, and Murray *et al.*, U.S. Patent No. 4,845,075. Transformed cells are selected by phenotype determined by the selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient (*e.g.*, leucine). An illustrative vector system for use in *Saccharomyces cerevisiae* is the *POT1* vector system disclosed by Kawasaki *et al.* (U.S. Patent No. 4,931,373), which allows transformed cells to be selected by growth in glucose-containing media. Additional suitable promoters and terminators for use in yeast include those from glycolytic enzyme genes (see, *e.g.*, Kawasaki, U.S. Patent No. 4,599,311, Kingsman *et al.*, U.S. Patent No. 4,615,974, and Bitter, U.S. Patent No. 4,977,092) and alcohol dehydrogenase genes. See also U.S. Patents Nos. 4,990,446, 5,063,154, 5,139,936, and 4,661,454.

Transformation systems for other yeasts, including *Hansenula polymorpha*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Ustilago maydis*, *Pichia pastoris*, *Pichia methanolica*, *Pichia guillermondii* and *Candida maltosa* are known in the art. See, for example, Gleeson *et al.*, *J. Gen. Microbiol.* 132:3459 (1986), and Cregg, U.S. Patent No. 4,882,279. *Aspergillus* cells may be utilized according to the methods of McKnight *et al.*, U.S. Patent No. 4,935,349. Methods for transforming *Acremonium chrysogenum* are disclosed by Sumino *et al.*,



U.S. Patent No. 5,162,228. Methods for transforming *Neurospora* are disclosed by Lambowitz, U.S. Patent No. 4,486,533.

For example, the use of *Pichia methanolica* as host for the production of recombinant proteins is disclosed by Raymond, U.S. Patent No. 5,716,808, Raymond, U.S. Patent No. 5,736,383, Raymond *et al.*, *Yeast* 14:11-23 (1998), and in international publication Nos. WO 97/17450, WO 97/17451, WO 98/02536, and WO 98/02565. DNA molecules for use in transforming *P. methanolica* will commonly be prepared as double-stranded, circular plasmids, which can be linearized prior to transformation. For polypeptide production in *P. methanolica*, the promoter and terminator in the plasmid can be that of a *P. methanolica* gene, such as a *P. methanolica* alcohol utilization gene (*AUG1* or *AUG2*). Other useful promoters include those of the dihydroxyacetone synthase (DHAS), formate dehydrogenase (FMD), and catalase (CAT) genes. To facilitate integration of the DNA into the host chromosome, the entire expression segment of the plasmid can be flanked at both ends by host DNA sequences. An illustrative selectable marker for use in *Pichia methanolica* is a *P. methanolica* *ADE2* gene, which encodes phosphoribosyl-5-aminoimidazole carboxylase (AIRC; EC 4.1.1.21), and which allows *ade2* host cells to grow in the absence of adenine. For large-scale, industrial processes where it is desirable to minimize the use of methanol, host cells can be used in which both methanol utilization genes (*AUG1* and *AUG2*) are deleted. For production of secreted proteins, host cells can be deficient in vacuolar protease genes (*PEP4* and *PRB1*). Electroporation is used to facilitate the introduction of a plasmid containing DNA encoding a polypeptide of interest into *P. methanolica* cells. *P. methanolica* cells can be transformed by electroporation using an exponentially decaying, pulsed electric field having a field strength of from 2.5 to 4.5 kV/cm, preferably about 3.75 kV/cm, and a time constant (t) of from 1 to 40 milliseconds, most preferably about 20 milliseconds.

Expression vectors can also be introduced into plant protoplasts, intact plant tissues, or isolated plant cells. Methods for introducing expression vectors into plant tissue include the direct infection or co-cultivation of plant tissue with *Agrobacterium tumefaciens*, microprojectile-mediated delivery, DNA injection, electroporation, and the like. See, for example, Horsch *et al.*, *Science* 227:1229 (1985), Klein *et al.*, *Biotechnology* 10:268 (1992), and Miki *et al.*, "Procedures for Introducing Foreign DNA into Plants," in *Methods in Plant Molecular Biology and Biotechnology*, Glick *et al.* (eds.), pages 67-88 (CRC Press, 1993).

Alternatively, *ZHMUP-2* genes can be expressed in prokaryotic host cells. Suitable promoters that can be used to express *ZHMUP-2* polypeptides in a prokaryotic host are well-known to those of skill in the art and include promoters capable of

recognizing the T4, T3, Sp6 and T7 polymerases, the P<sub>R</sub> and P<sub>L</sub> promoters of bacteriophage lambda, the *trp*, *recA*, heat shock, *lacUV5*, *tac*, *lpp-lacSpr*, *phoA*, and *lacZ* promoters of *E. coli*, promoters of *B. subtilis*, the promoters of the bacteriophages of *Bacillus*, *Streptomyces* promoters, the *int* promoter of bacteriophage lambda, the *bla* promoter of pBR322, and the CAT promoter of the chloramphenicol acetyl transferase gene. Prokaryotic promoters have been reviewed by Glick, *J. Ind. Microbiol.* 1:277 (1987), Watson *et al.*, *Molecular Biology of the Gene*, 4th Ed. (Benjamin Cummins 1987), and by Ausubel *et al.* (1995).

Useful prokaryotic hosts include *E. coli* and *Bacillus subtilis*. Suitable strains of *E. coli* include BL21(DE3), BL21(DE3)pLysS, BL21(DE3)pLysE, DH1, DH4I, DH5, DH5I, DH5IF', DH5IMCR, DH10B, DH10B/p3, DH11S, C600, HB101, JM101, JM105, JM109, JM110, K38, RR1, Y1088, Y1089, CSH18, ER1451, and ER1647 (see, for example, Brown (ed.), *Molecular Biology Labfax* (Academic Press 1991)). Suitable strains of *Bacillus subtilis* include BR151, YB886, MI119, MI120, and B170 (see, for example, Hardy, "Bacillus Cloning Methods," in *DNA Cloning: A Practical Approach*, Glover (ed.) (IRL Press 1985)).

When expressing a ZHMUP-2 polypeptide in bacteria such as *E. coli*, the polypeptide may be retained in the cytoplasm, typically as insoluble granules, or may be directed to the periplasmic space by a bacterial secretion sequence. In the former case, the cells are lysed, and the granules are recovered and denatured using, for example, guanidine isothiocyanate or urea. The denatured polypeptide can then be refolded and dimerized by diluting the denaturant, such as by dialysis against a solution of urea and a combination of reduced and oxidized glutathione, followed by dialysis against a buffered saline solution. In the latter case, the polypeptide can be recovered from the periplasmic space in a soluble and functional form by disrupting the cells (by, for example, sonication or osmotic shock) to release the contents of the periplasmic space and recovering the protein, thereby obviating the need for denaturation and refolding.

Methods for expressing proteins in prokaryotic hosts are well-known to those of skill in the art (see, for example, Williams *et al.*, "Expression of foreign proteins in *E. coli* using plasmid vectors and purification of specific polyclonal antibodies," in *DNA Cloning 2: Expression Systems*, 2nd Edition, Glover *et al.* (eds.), page 15 (Oxford University Press 1995), Ward *et al.*, "Genetic Manipulation and Expression of Antibodies," in *Monoclonal Antibodies: Principles and Applications*, page 137 (Wiley-Liss, Inc. 1995), and Georgiou, "Expression of Proteins in Bacteria," in *Protein Engineering: Principles and Practice*, Cleland *et al.* (eds.), page 101 (John Wiley & Sons, Inc. 1996)).

Standard methods for introducing expression vectors into bacterial, yeast, insect, and plant cells are provided, for example, by Ausubel (1995).

General methods for expressing and recovering foreign protein produced by a mammalian cell system are provided by, for example, Etcheverry, "Expression of Engineered Proteins in Mammalian Cell Culture," in *Protein Engineering: Principles and Practice*, Cleland *et al.* (eds.), pages 163 (Wiley-Liss, Inc. 1996). Standard techniques for recovering protein produced by a bacterial system is provided by, for example, Grisshammer *et al.*, "Purification of over-produced proteins from *E. coli* cells," in *DNA Cloning 2: Expression Systems, 2nd Edition*, Glover *et al.* (eds.), pages 59-92 (Oxford University Press 1995). Established methods for isolating recombinant proteins from a baculovirus system are described by Richardson (ed.), *Baculovirus Expression Protocols* (The Humana Press, Inc. 1995). In addition, production of functional murine MUP has been demonstrated in *Pichia pastoris* and in *E. coli* (Ferrari *et al.*, *FEBS Lett.* 401:73 (1997); Zidek *et al.*, *Biochemistry* 38:9850 (1999)).

As an alternative, polypeptides of the present invention can be synthesized by exclusive solid phase synthesis, partial solid phase methods, fragment condensation or classical solution synthesis. These synthesis methods are well-known to those of skill in the art (see, for example, Merrifield, *J. Am. Chem. Soc.* 85:2149 (1963), Stewart *et al.*, "Solid Phase Peptide Synthesis" (2nd Edition), (Pierce Chemical Co. 1984), Bayer and Rapp, *Chem. Pept. Prot.* 3:3 (1986), Atherton *et al.*, *Solid Phase Peptide Synthesis: A Practical Approach* (IRL Press 1989), Fields and Colowick, "Solid-Phase Peptide Synthesis," *Methods in Enzymology Volume 289* (Academic Press 1997), and Lloyd-Williams *et al.*, *Chemical Approaches to the Synthesis of Peptides and Proteins* (CRC Press, Inc. 1997)). Variations in total chemical synthesis strategies, such as "native chemical ligation" and "expressed protein ligation" are also standard (see, for example, Dawson *et al.*, *Science* 266:776 (1994), Hackeng *et al.*, *Proc. Nat'l Acad. Sci. USA* 94:7845 (1997), Dawson, *Methods Enzymol.* 287: 34 (1997), Muir *et al.*, *Proc. Nat'l Acad. Sci. USA* 95:6705 (1998), Severinov and Muir, *J. Biol. Chem.* 273:16205 (1998), and Dawson and Kent, *Annu. Rev. Biochem.* 69:923 (2000)).

## 7. Isolation of ZHMUP-2 Polypeptides

The polypeptides of the present invention can be purified to at least about 80% purity, to at least about 90% purity, to at least about 95% purity, or greater than 95% purity with respect to contaminating macromolecules, particularly other proteins and nucleic acids, and free of infectious and pyrogenic agents. The polypeptides of the present invention may also be purified to a pharmaceutically pure state, which is greater

than 99.9% pure. Certain purified polypeptide preparations are substantially free of other polypeptides, particularly other polypeptides of animal origin.

Fractionation and/or conventional purification methods can be used to obtain preparations of ZHMUP-2 purified from natural sources, and recombinant ZHMUP-2 polypeptides and fusion ZHMUP-2 polypeptides purified from recombinant host cells. In general, ammonium sulfate precipitation and acid or chaotrope extraction may be used for fractionation of samples. Exemplary purification steps may include hydroxyapatite, size exclusion, FPLC and reverse-phase high performance liquid chromatography. Suitable chromatographic media include derivatized dextrans, agarose, cellulose, polyacrylamide, specialty silicas, and the like. PEI, DEAE, QAE and Q derivatives are preferred. Exemplary chromatographic media include those media derivatized with phenyl, butyl, or octyl groups, such as Phenyl-Sepharose FF (Pharmacia), Toyopearl butyl 650 (Toso Haas, Montgomeryville, PA), Octyl-Sepharose (Pharmacia) and the like; or polyacrylic resins, such as Amberchrom CG 71 (Toso Haas) and the like. Suitable solid supports include glass beads, silica-based resins, cellulosic resins, agarose beads, cross-linked agarose beads, polystyrene beads, cross-linked polyacrylamide resins and the like that are insoluble under the conditions in which they are to be used. These supports may be modified with reactive groups that allow attachment of proteins by amino groups, carboxyl groups, sulfhydryl groups, hydroxyl groups and/or carbohydrate moieties.

Examples of coupling chemistries include cyanogen bromide activation, N-hydroxysuccinimide activation, epoxide activation, sulfhydryl activation, hydrazide activation, and carboxyl and amino derivatives for carbodiimide coupling chemistries. These and other solid media are well known and widely used in the art, and are available from commercial suppliers. Selection of a particular method for polypeptide isolation and purification is a matter of routine design and is determined in part by the properties of the chosen support. See, for example, *Affinity Chromatography: Principles & Methods* (Pharmacia LKB Biotechnology 1988), and Doonan, *Protein Purification Protocols* (The Humana Press 1996).

Additional variations in ZHMUP-2 isolation and purification can be devised by those of skill in the art. For example, anti-ZHMUP-2 antibodies, obtained as described below, can be used to isolate large quantities of protein by immunoaffinity purification. Moreover, methods for binding ligands, such as ZHMUP-2, to receptor polypeptides bound to support media are well known in the art.

The polypeptides of the present invention can also be isolated by exploitation of particular properties. For example, immobilized metal ion adsorption (IMAC) chromatography can be used to purify histidine-rich proteins, including those

comprising polyhistidine tags. Briefly, a gel is first charged with divalent metal ions to form a chelate (Sulkowski, *Trends in Biochem.* 3:1 (1985)). Histidine-rich proteins will be adsorbed to this matrix with differing affinities, depending upon the metal ion used, and will be eluted by competitive elution, lowering the pH, or use of strong chelating agents. Other methods of purification include purification of glycosylated proteins by lectin affinity chromatography and ion exchange chromatography (M. Deutscher, (ed.), *Meth. Enzymol.* 182:529 (1990)). Within additional embodiments of the invention, a fusion of the polypeptide of interest and an affinity tag (e.g., maltose-binding protein, an immunoglobulin domain) may be constructed to facilitate purification.

ZHMUP-2 polypeptides or fragments thereof may also be prepared through chemical synthesis, as described above. ZHMUP-2 polypeptides may be monomers or multimers; glycosylated or non-glycosylated; and may or may not include an initial methionine amino acid residue.

The present invention also contemplates chemically modified ZHMUP-2 compositions, in which a ZHMUP-2 polypeptide is linked with a polymer. Typically, the polymer is water soluble so that the ZHMUP-2 conjugate does not precipitate in an aqueous environment, such as a physiological environment. An example of a suitable polymer is one that has been modified to have a single reactive group, such as an active ester for acylation, or an aldehyde for alkylation. In this way, the degree of polymerization can be controlled. An example of a reactive aldehyde is polyethylene glycol propionaldehyde, or mono-(C<sub>1</sub>-C<sub>10</sub>) alkoxy, or aryloxy derivatives thereof (see, for example, Harris, *et al.*, U.S. Patent No. 5,252,714). The polymer may be branched or unbranched. Moreover, a mixture of polymers can be used to produce ZHMUP-2 conjugates.

ZHMUP-2 conjugates used for therapy should can comprise pharmaceutically acceptable water-soluble polymer moieties. Suitable water-soluble polymers include polyethylene glycol (PEG), monomethoxy-PEG, mono-(C<sub>1</sub>-C<sub>10</sub>)alkoxy-PEG, aryloxy-PEG, poly-(N-vinyl pyrrolidone)PEG, tresyl monomethoxy PEG, PEG propionaldehyde, bis-succinimidyl carbonate PEG, propylene glycol homopolymers, a polypropylene oxide/ethylene oxide co-polymer, polyoxyethylated polyols (e.g., glycerol), polyvinyl alcohol, dextran, cellulose, or other carbohydrate-based polymers. Suitable PEG may have a molecular weight from about 600 to about 60,000, including, for example, 5,000, 12,000, 20,000 and 25,000. A ZHMUP-2 conjugate can also comprise a mixture of such water-soluble polymers. Anti-ZHMUP-2 antibodies or anti-idiotypic antibodies can also be conjugated with a water-soluble polymer.

The present invention contemplates compositions comprising a peptide or polypeptide described herein. Such compositions can further comprise a carrier. The

carrier can be a conventional organic or inorganic carrier. Examples of carriers include water, buffer solution, alcohol, propylene glycol, macrogol, sesame oil, corn oil, and the like.

Peptides and polypeptides of the present invention comprise at least six, at least nine, or at least 15 contiguous amino acid residues of SEQ ID NO:2, the amino acid sequence of amino acid residues 46 to 185 of SEQ ID NO:2, the amino acid sequence of amino acid residues 16 to 185 of SEQ ID NO:2, the amino acid sequence of amino acid residues 1 to 32 of SEQ ID NO:2, the amino acid sequence of amino acid residues 16 to 32 of SEQ ID NO:2, the amino acid sequence of amino acid residues 33 to 77 of SEQ ID NO:2, the amino acid sequence of amino acid residues 79 to 102 of SEQ ID NO:2, the amino acid sequence of amino acid residues 104 to 139 of SEQ ID NO:2, the amino acid sequence of amino acid residues 141 to 173 of SEQ ID NO:2, and the amino acid sequence of amino acid residues 175 to 185 of SEQ ID NO:2. Within certain embodiments of the invention, the polypeptides comprise 20, 30, 40, 50, 100, or more contiguous residues of these amino acid sequences. Nucleic acid molecules encoding such peptides and polypeptides are useful as polymerase chain reaction primers and probes, and these peptides and polypeptides are useful to produce antibodies to ZHMUP-2.

## 8. *Production of Antibodies to ZHMUP-2 Proteins*

Antibodies to ZHMUP-2 can be obtained, for example, using as an antigen the product of a ZHMUP-2 expression vector or ZHMUP-2 isolated from a natural source. Particularly useful anti-ZHMUP-2 antibodies "bind specifically" with ZHMUP-2. Antibodies are considered to be specifically binding if the antibodies exhibit at least one of the following two properties: (1) antibodies bind to ZHMUP-2 with a threshold level of binding activity, and (2) antibodies do not significantly cross-react with polypeptides related to ZHMUP-2, such as known murine major urinary proteins, and porcine sex-specific salivary lipocalin.

With regard to the first characteristic, antibodies specifically bind if they bind to a ZHMUP-2 polypeptide, peptide or epitope with a binding affinity ( $K_a$ ) of  $10^6 M^{-1}$  or greater, preferably  $10^7 M^{-1}$  or greater, more preferably  $10^8 M^{-1}$  or greater, and most preferably  $10^9 M^{-1}$  or greater. The binding affinity of an antibody can be readily determined by one of ordinary skill in the art, for example, by Scatchard analysis (Scatchard, *Ann. NY Acad. Sci.* 51:660 (1949)). With regard to the second characteristic, antibodies do not significantly cross-react with related polypeptide molecules, for

example, if they detect ZHMUP-2, but not known polypeptides, using a standard Western blot analysis.

Anti-ZHMUP-2 antibodies can be produced using antigenic ZHMUP-2 epitope-bearing peptides and polypeptides. Antigenic epitope-bearing peptides and polypeptides of the present invention contain a sequence of at least nine, or between 15 to about 30 amino acids contained within SEQ ID NO:2. However, peptides or polypeptides comprising a larger portion of an amino acid sequence of the invention, containing from 30 to 50 amino acids, or any length up to and including the entire amino acid sequence of a polypeptide of the invention, also are useful for inducing antibodies that bind with ZHMUP-2. It is desirable that the amino acid sequence of the epitope-bearing peptide is selected to provide substantial solubility in aqueous solvents (*i.e.*, the sequence includes relatively hydrophilic residues, while hydrophobic residues are preferably avoided). Moreover, amino acid sequences containing proline residues may be also be desirable for antibody production.

As an illustration, potential antigenic sites in ZHMUP-2 were identified using the Jameson-Wolf method, Jameson and Wolf, *CABIOS* 4:181, (1988), as implemented by the PROTEAN program (version 3.14) of LASERGENE (DNASTAR; Madison, WI). Default parameters were used in this analysis.

The Jameson-Wolf method predicts potential antigenic determinants by combining six major subroutines for protein structural prediction. Briefly, the Hopp-Woods method, Hopp *et al.*, *Proc. Nat'l Acad. Sci. USA* 78:3824 (1981), was first used to identify amino acid sequences representing areas of greatest local hydrophilicity (parameter: seven residues averaged). In the second step, Emini's method, Emini *et al.*, *J. Virology* 55:836 (1985), was used to calculate surface probabilities (parameter: surface decision threshold (0.6) = 1). Third, the Karplus-Schultz method, Karplus and Schultz, *Naturwissenschaften* 72:212 (1985), was used to predict backbone chain flexibility (parameter: flexibility threshold (0.2) = 1). In the fourth and fifth steps of the analysis, secondary structure predictions were applied to the data using the methods of Chou-Fasman, Chou, "Prediction of Protein Structural Classes from Amino Acid Composition," in *Prediction of Protein Structure and the Principles of Protein Conformation*, Fasman (ed.), pages 549-586 (Plenum Press 1990), and Garnier-Robson, Garnier *et al.*, *J. Mol. Biol.* 120:97 (1978) (Chou-Fasman parameters: conformation table = 64 proteins;  $\alpha$  region threshold = 103;  $\beta$  region threshold = 105; Garnier-Robson parameters:  $\alpha$  and  $\beta$  decision constants = 0). In the sixth subroutine, flexibility parameters and hydropathy/solvent accessibility factors were combined to determine a surface contour value, designated as the "antigenic index." Finally, a peak broadening function was applied to the antigenic index, which broadens major surface peaks by

adding 20, 40, 60, or 80% of the respective peak value to account for additional free energy derived from the mobility of surface regions relative to interior regions. This calculation was not applied, however, to any major peak that resides in a helical region, since helical regions tend to be less flexible.

5           The results of this analysis indicated that the following illustrative amino acid sequences of SEQ ID NO:2 would provide suitable antigenic molecules: amino acid residues 16 to 28 ("antigenic molecule 1"), amino acid residues 31 to 37 ("antigenic molecule 2"), amino acid residues 44 to 56 ("antigenic molecule 3"), amino acid residues 78 to 86 ("antigenic molecule 4"), amino acid residues 91 to 98 ("antigenic molecule 5"),  
10       amino acid residues 112 to 119 ("antigenic molecule 6"), amino acid residues 125 to 132 ("antigenic molecule 7"), amino acid residues 142 to 154 ("antigenic molecule 8"), amino acid residues 161 to 167 ("antigenic molecule 9"), and amino acid residues 171 to 184 ("antigenic molecule 10"). The present invention contemplates the use of any one of antigenic molecules 1 to 10 to generate antibodies to ZHMUP-2 proteins. The present  
15       invention also contemplates polypeptides comprising at least one of antigenic molecules 1 to 10.

          Useful antibodies can also be produced using antigenic molecules that comprise at least one *ZHMUP-2* exon. For example, such antigenic molecules can comprise polypeptides that consist of the following amino acid sequences of SEQ ID  
20       NO:2: amino acid residues 1 to 32, amino acid residues 33 to 77, amino acid residues 79 to 102, amino acid residues 104 to 139, amino acid residues 141 to 173, and amino acid residues 175 to amino acid residue 185.

          Polyclonal antibodies to recombinant *ZHMUP-2* protein or to *ZHMUP-2* isolated from natural sources can be prepared using methods well-known to those of skill  
25       in the art. Antibodies can also be generated using a *ZHMUP-2*-glutathione transferase fusion protein, which is similar to a method described by Burrus and McMahon, *Exp. Cell. Res.* 220:363 (1995). General methods for producing polyclonal antibodies are described, for example, by Green *et al.*, "Production of Polyclonal Antisera," in *Immunochemical Protocols* (Manson, ed.), pages 1-5 (Humana Press 1992), and  
30       Williams *et al.*, "Expression of foreign proteins in *E. coli* using plasmid vectors and purification of specific polyclonal antibodies," in *DNA Cloning 2: Expression Systems, 2nd Edition*, Glover *et al.* (eds.), page 15 (Oxford University Press 1995).

          The immunogenicity of a *ZHMUP-2* polypeptide can be increased through the use of an adjuvant, such as alum (aluminum hydroxide) or Freund's complete  
35       or incomplete adjuvant. Polypeptides useful for immunization also include fusion polypeptides, such as fusions of *ZHMUP-2* or a portion thereof with an immunoglobulin polypeptide or with maltose binding protein. The polypeptide immunogen may be a full-



length molecule or a portion thereof. If the polypeptide portion is "hapten-like," such portion may be advantageously joined or linked to a macromolecular carrier (such as keyhole limpet hemocyanin (KLH), bovine serum albumin (BSA) or tetanus toxoid) for immunization.

5           Although polyclonal antibodies are typically raised in animals such as horse, cow, dog, chicken, rat, mouse, rabbit, goat, guinea pig, or sheep, an anti-ZHMUP-2 antibody of the present invention may also be derived from a subhuman primate antibody. General techniques for raising diagnostically and therapeutically useful antibodies in baboons may be found, for example, in Goldenberg *et al.*, international  
10 patent publication No. WO 91/11465, and in Losman *et al.*, *Int. J. Cancer* 46:310 (1990).

          Alternatively, monoclonal anti-ZHMUP-2 antibodies can be generated. Rodent monoclonal antibodies to specific antigens may be obtained by methods known to those skilled in the art (see, for example, Kohler *et al.*, *Nature* 256:495 (1975),  
15 Coligan *et al.* (eds.), *Current Protocols in Immunology*, Vol. 1, pages 2.5.1-2.6.7 (John Wiley & Sons 1991) ["Coligan"], Picksley *et al.*, "Production of monoclonal antibodies against proteins expressed in *E. coli*," in *DNA Cloning 2: Expression Systems*, 2nd Edition, Glover *et al.* (eds.), page 93 (Oxford University Press 1995)).

          Briefly, monoclonal antibodies can be obtained by injecting mice with a  
20 composition comprising a ZHMUP-2 gene product, verifying the presence of antibody production by removing a serum sample, removing the spleen to obtain B-lymphocytes, fusing the B-lymphocytes with myeloma cells to produce hybridomas, cloning the hybridomas, selecting positive clones which produce antibodies to the antigen, culturing the clones that produce antibodies to the antigen, and isolating the antibodies from the  
25 hybridoma cultures.

          In addition, an anti-ZHMUP-2 antibody of the present invention may be derived from a human monoclonal antibody. Human monoclonal antibodies are obtained from transgenic mice that have been engineered to produce specific human antibodies in response to antigenic challenge. In this technique, elements of the human heavy and light  
30 chain locus are introduced into strains of mice derived from embryonic stem cell lines that contain targeted disruptions of the endogenous heavy chain and light chain loci. The transgenic mice can synthesize human antibodies specific for human antigens, and the mice can be used to produce human antibody-secreting hybridomas. Methods for obtaining human antibodies from transgenic mice are described, for example, by Green *et al.*, *Nature Genet.* 7:13 (1994), Lonberg *et al.*, *Nature* 368:856 (1994), and Taylor *et al.*, *Int. Immun.* 6:579 (1994).  
35

Monoclonal antibodies can be isolated and purified from hybridoma cultures by a variety of well-established techniques. Such isolation techniques include affinity chromatography with Protein-A Sepharose, size-exclusion chromatography, and ion-exchange chromatography (see, for example, Coligan at pages 2.7.1-2.7.12 and pages 2.9.1-2.9.3; Baines *et al.*, "Purification of Immunoglobulin G (IgG)," in *Methods in Molecular Biology*, Vol. 10, pages 79-104 (The Humana Press, Inc. 1992)).

For particular uses, it may be desirable to prepare fragments of anti-ZHMUP-2 antibodies. Such antibody fragments can be obtained, for example, by proteolytic hydrolysis of the antibody. Antibody fragments can be obtained by pepsin or papain digestion of whole antibodies by conventional methods. As an illustration, antibody fragments can be produced by enzymatic cleavage of antibodies with pepsin to provide a 5S fragment denoted F(ab')<sub>2</sub>. This fragment can be further cleaved using a thiol reducing agent to produce 3.5S Fab' monovalent fragments. Optionally, the cleavage reaction can be performed using a blocking group for the sulfhydryl groups that result from cleavage of disulfide linkages. As an alternative, an enzymatic cleavage using pepsin produces two monovalent Fab fragments and an Fc fragment directly. These methods are described, for example, by Goldenberg, U.S. patent No. 4,331,647, Nisonoff *et al.*, *Arch Biochem. Biophys.* 89:230 (1960), Porter, *Biochem. J.* 73:119 (1959), Edelman *et al.*, in *Methods in Enzymology Vol. 1*, page 422 (Academic Press 1967), and by Coligan at pages 2.8.1-2.8.10 and 2.10.-2.10.4.

Other methods of cleaving antibodies, such as separation of heavy chains to form monovalent light-heavy chain fragments, further cleavage of fragments, or other enzymatic, chemical or genetic techniques may also be used, so long as the fragments bind to the antigen that is recognized by the intact antibody.

For example, Fv fragments comprise an association of V<sub>H</sub> and V<sub>L</sub> chains. This association can be noncovalent, as described by Inbar *et al.*, *Proc. Nat'l Acad. Sci. USA* 69:2659 (1972). Alternatively, the variable chains can be linked by an intermolecular disulfide bond or cross-linked by chemicals such as glutaraldehyde (see, for example, Sandhu, *Crit. Rev. Biotech.* 12:437 (1992)).

The Fv fragments may comprise V<sub>H</sub> and V<sub>L</sub> chains which are connected by a peptide linker. These single-chain antigen binding proteins (scFv) are prepared by constructing a structural gene comprising DNA sequences encoding the V<sub>H</sub> and V<sub>L</sub> domains which are connected by an oligonucleotide. The structural gene is inserted into an expression vector which is subsequently introduced into a host cell, such as *E. coli*. The recombinant host cells synthesize a single polypeptide chain with a linker peptide bridging the two V domains. Methods for producing scFvs are described, for example, by Whitlow *et al.*, *Methods: A Companion to Methods in Enzymology* 2:97 (1991) (also

see, Bird *et al.*, *Science* 242:423 (1988), Ladner *et al.*, U.S. Patent No. 4,946,778, Pack *et al.*, *Bio/Technology* 11:1271 (1993), and Sandhu, *supra*).

As an illustration, an scFV can be obtained by exposing lymphocytes to ZHMUP-2 polypeptide *in vitro*, and selecting antibody display libraries in phage or similar vectors (for instance, through use of immobilized or labeled ZHMUP-2 protein or peptide). Genes encoding polypeptides having potential ZHMUP-2 polypeptide binding domains can be obtained by screening random peptide libraries displayed on phage (phage display) or on bacteria, such as *E. coli*. Nucleotide sequences encoding the polypeptides can be obtained in a number of ways, such as through random mutagenesis and random polynucleotide synthesis. These random peptide display libraries can be used to screen for peptides, which interact with a known target which can be a protein or polypeptide, such as a ligand or receptor, a biological or synthetic macromolecule, or organic or inorganic substances. Techniques for creating and screening such random peptide display libraries are known in the art (Ladner *et al.*, U.S. Patent No. 5,223,409, Ladner *et al.*, U.S. Patent No. 4,946,778, Ladner *et al.*, U.S. Patent No. 5,403,484, Ladner *et al.*, U.S. Patent No. 5,571,698, and Kay *et al.*, *Phage Display of Peptides and Proteins* (Academic Press, Inc. 1996)) and random peptide display libraries and kits for screening such libraries are available commercially, for instance from CLONTECH Laboratories, Inc. (Palo Alto, CA), Invitrogen Inc. (San Diego, CA), New England Biolabs, Inc. (Beverly, MA), and Pharmacia LKB Biotechnology Inc. (Piscataway, NJ). Random peptide display libraries can be screened using the ZHMUP-2 sequences disclosed herein to identify proteins which bind to ZHMUP-2.

Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. Such genes are prepared, for example, by using the polymerase chain reaction to synthesize the variable region from RNA of antibody-producing cells (see, for example, Larrick *et al.*, *Methods: A Companion to Methods in Enzymology* 2:106 (1991), Courtenay-Luck, "Genetic Manipulation of Monoclonal Antibodies," in *Monoclonal Antibodies: Production, Engineering and Clinical Application*, Ritter *et al.* (eds.), page 166 (Cambridge University Press 1995), and Ward *et al.*, "Genetic Manipulation and Expression of Antibodies," in *Monoclonal Antibodies: Principles and Applications*, Birch *et al.*, (eds.), page 137 (Wiley-Liss, Inc. 1995)).

Alternatively, an anti-ZHMUP-2 antibody may be derived from a "humanized" monoclonal antibody. Humanized monoclonal antibodies are produced by transferring mouse complementary determining regions from heavy and light variable chains of the mouse immunoglobulin into a human variable domain. Typical residues of

human antibodies are then substituted in the framework regions of the murine counterparts. The use of antibody components derived from humanized monoclonal antibodies obviates potential problems associated with the immunogenicity of murine constant regions. General techniques for cloning murine immunoglobulin variable domains are described, for example, by Orlandi *et al.*, *Proc. Nat'l Acad. Sci. USA* 86:3833 (1989). Techniques for producing humanized monoclonal antibodies are described, for example, by Jones *et al.*, *Nature* 321:522 (1986), Carter *et al.*, *Proc. Nat'l Acad. Sci. USA* 89:4285 (1992), Sandhu, *Crit. Rev. Biotech.* 12:437 (1992), Singer *et al.*, *J. Immun.* 150:2844 (1993), Sudhir (ed.), *Antibody Engineering Protocols* (Humana Press, Inc. 1995), Kelley, "Engineering Therapeutic Antibodies," in *Protein Engineering: Principles and Practice*, Cleland *et al.* (eds.), pages 399-434 (John Wiley & Sons, Inc. 1996), and by Queen *et al.*, U.S. Patent No. 5,693,762 (1997).

Polyclonal anti-idiotypic antibodies can be prepared by immunizing animals with anti-ZHMUP-2 antibodies or antibody fragments, using standard techniques. See, for example, Green *et al.*, "Production of Polyclonal Antisera," in *Methods In Molecular Biology: Immunochemical Protocols*, Manson (ed.), pages 1-12 (Humana Press 1992). Also, see Coligan at pages 2.4.1-2.4.7. Alternatively, monoclonal anti-idiotypic antibodies can be prepared using anti-ZHMUP-2 antibodies or antibody fragments as immunogens with the techniques, described above. As another alternative, humanized anti-idiotypic antibodies or subhuman primate anti-idiotypic antibodies can be prepared using the above-described techniques. Methods for producing anti-idiotypic antibodies are described, for example, by Irie, U.S. Patent No. 5,208,146, Greene, *et al.*, U.S. Patent No. 5,637,677, and Varthakavi and Minocha, *J. Gen. Virol.* 77:1875 (1996).

#### **9. Use of ZHMUP-2 Nucleotide Sequences to Detect ZHMUP-2 Gene Expression and to Examine ZHMUP-2 Gene Structure**

Nucleic acid molecules can be used to detect the expression of a ZHMUP-2 gene in a biological sample. Such probe molecules include double-stranded nucleic acid molecules comprising the nucleotide sequence of SEQ ID NO:1, or a portion thereof, as well as single-stranded nucleic acid molecules having the complement of the nucleotide sequence of SEQ ID NO:1, or a portion thereof. As used herein, the term "portion" refers to at least eight nucleotides to at least 20 or more nucleotides. Probe molecules may be DNA, RNA, oligonucleotides, and the like.

In a basic assay, a single-stranded probe molecule is incubated with RNA, isolated from a biological sample, under conditions of temperature and ionic strength that promote base pairing between the probe and target ZHMUP-2 RNA species. After separating unbound probe from hybridized molecules, the amount of hybrids is detected.

5 Well-established hybridization methods of RNA detection include northern analysis and dot/slot blot hybridization (see, for example, Ausubel (1995) at pages 4-1 to 4-27, and Wu *et al.* (eds.), "Analysis of Gene Expression at the RNA Level," in *Methods in Gene Biotechnology*, pages 225-239 (CRC Press, Inc. 1997)). Nucleic acid probes can be detectably labeled with radioisotopes such as  $^{32}\text{P}$  or  $^{35}\text{S}$ .  
10 Alternatively, ZHMUP-2 RNA can be detected with a nonradioactive hybridization method (see, for example, Isaac (ed.), *Protocols for Nucleic Acid Analysis by Nonradioactive Probes* (Humana Press, Inc. 1993)). Typically, nonradioactive detection is achieved by enzymatic conversion of chromogenic or chemiluminescent substrates. Illustrative nonradioactive moieties include biotin, fluorescein, and digoxigenin.

15 ZHMUP-2 oligonucleotide probes are also useful for *in vivo* diagnosis. As an illustration,  $^{18}\text{F}$ -labeled oligonucleotides can be administered to a subject and visualized by positron emission tomography (Tavittian *et al.*, *Nature Medicine* 4:467 (1998)).

Numerous diagnostic procedures take advantage of the polymerase chain reaction (PCR) to increase sensitivity of detection methods. Standard techniques for  
20 performing PCR are well-known (see, generally, Mathew (ed.), *Protocols in Human Molecular Genetics* (Humana Press, Inc. 1991), White (ed.), *PCR Protocols: Current Methods and Applications* (Humana Press, Inc. 1993), Cotter (ed.), *Molecular Diagnosis of Cancer* (Humana Press, Inc. 1996), Hanausek and Walaszek (eds.), *Tumor Marker Protocols* (Humana Press, Inc. 1998), Lo (ed.), *Clinical Applications of PCR* (Humana Press, Inc. 1998), and Meltzer (ed.), *PCR in Bioanalysis* (Humana Press, Inc. 1998)).

25 One variation of PCR for diagnostic assays is reverse transcriptase-PCR (RT-PCR). In the RT-PCR technique, RNA is isolated from a biological sample, reverse transcribed to cDNA, and the cDNA is incubated with ZHMUP-2 primers (see, for example, Wu *et al.* (eds.), "Rapid Isolation of Specific cDNAs or Genes by PCR," in  
30 *Methods in Gene Biotechnology*, pages 15-28 (CRC Press, Inc. 1997)). PCR is then performed and the products are analyzed using standard techniques.

As an illustration, RNA is isolated from biological sample using, for example, the guanidinium-thiocyanate cell lysis procedure described above. Alternatively, a solid-phase technique can be used to isolate mRNA from a cell lysate. A  
35 reverse transcription reaction can be primed with the isolated RNA using random oligonucleotides, short homopolymers of dT, or ZHMUP-2 anti-sense oligomers. Oligo-dT primers offer the advantage that various mRNA nucleotide sequences are amplified

that can provide control target sequences. ZHMUP-2 sequences are amplified by the polymerase chain reaction using two flanking oligonucleotide primers that are typically 20 bases in length.

5 PCR amplification products can be detected using a variety of approaches. For example, PCR products can be fractionated by gel electrophoresis, and visualized by ethidium bromide staining. Alternatively, fractionated PCR products can be transferred to a membrane, hybridized with a detectably-labeled ZHMUP-2 probe, and examined by autoradiography. Additional alternative approaches include the use of digoxigenin-labeled deoxyribonucleic acid triphosphates to provide chemiluminescence  
10 detection, and the C-TRAK colorimetric assay.

Another approach for detection of ZHMUP-2 expression is cycling probe technology (CPT), in which a single-stranded DNA target binds with an excess of DNA-RNA-DNA chimeric probe to form a complex, the RNA portion is cleaved with RNAase H, and the presence of cleaved chimeric probe is detected (see, for example, Beggs *et al.*,  
15 *J. Clin. Microbiol.* 34:2985 (1996), Bekkaoui *et al.*, *Biotechniques* 20:240 (1996)). Alternative methods for detection of ZHMUP-2 sequences can utilize approaches such as nucleic acid sequence-based amplification (NASBA), cooperative amplification of templates by cross-hybridization (CATCH), and the ligase chain reaction (LCR) (see, for example, Marshall *et al.*, U.S. Patent No. 5,686,272 (1997), Dyer *et al.*, *J. Virol.*  
20 *Methods* 60:161 (1996), Ehricht *et al.*, *Eur. J. Biochem.* 243:358 (1997), and Chadwick *et al.*, *J. Virol. Methods* 70:59 (1998)). Other standard methods are known to those of skill in the art.

ZHMUP-2 probes and primers can also be used to detect and to localize ZHMUP-2 gene expression in tissue samples. Methods for such *in situ* hybridization are  
25 well-known to those of skill in the art (see, for example, Choo (ed.), *In Situ Hybridization Protocols* (Humana Press, Inc. 1994), Wu *et al.* (eds.), "Analysis of Cellular DNA or Abundance of mRNA by Radioactive *In Situ* Hybridization (RISH)," in *Methods in Gene Biotechnology*, pages 259-278 (CRC Press, Inc. 1997), and Wu *et al.* (eds.), "Localization of DNA or Abundance of mRNA by Fluorescence *In Situ* Hybridization (RISH)," in  
30 *Methods in Gene Biotechnology*, pages 279-289 (CRC Press, Inc. 1997)). Various additional diagnostic approaches are well-known to those of skill in the art (see, for example, Mathew (ed.), *Protocols in Human Molecular Genetics* (Humana Press, Inc. 1991), Coleman and Tsongalis, *Molecular Diagnostics* (Humana Press, Inc. 1996), and Elles, *Molecular Diagnosis of Genetic Diseases* (Humana Press, Inc., 1996)). Suitable  
35 test samples include blood, urine, saliva, tissue biopsy, and autopsy material.

The ZHMUP-2 gene resides in chromosome 6p11.2 - 21, a region that is associated with diseases and disorders, such as retinal degeneration, and nystagmus.

Nucleic acid molecules comprising *ZHMUP-2* nucleotide sequences can be used in linkage-based testing for various diseases, and to determine whether a subject's chromosomes contain a mutation in the *ZHMUP-2* gene. Of particular interest are genetic alterations that inactivate a *ZHMUP-2* gene.

5 Aberrations associated with a *ZHMUP-2* locus can be detected using nucleic acid molecules of the present invention by employing molecular genetic techniques, such as restriction fragment length polymorphism analysis, short tandem repeat analysis employing PCR techniques, amplification-refractory mutation system analysis, single-strand conformation polymorphism detection, RNase cleavage methods, 10 denaturing gradient gel electrophoresis, fluorescence-assisted mismatch analysis, and other genetic analysis techniques known in the art (see, for example, Mathew (ed.), *Protocols in Human Molecular Genetics* (Humana Press, Inc. 1991), Marian, *Chest* 108:255 (1995), Coleman and Tsongalis, *Molecular Diagnostics* (Human Press, Inc. 1996), Elles (ed.) *Molecular Diagnosis of Genetic Diseases* (Humana Press, Inc. 1996), 15 Landegren (ed.), *Laboratory Protocols for Mutation Detection* (Oxford University Press 1996), Birren *et al.* (eds.), *Genome Analysis, Vol. 2: Detecting Genes* (Cold Spring Harbor Laboratory Press 1998), Dracopoli *et al.* (eds.), *Current Protocols in Human Genetics* (John Wiley & Sons 1998), and Richards and Ward, "Molecular Diagnostic Testing," in *Principles of Molecular Medicine*, pages 83-88 (Humana Press, Inc. 1998)).

20 As an illustration, large deletions in a *ZHMUP-2* gene can be detected using Southern hybridization analysis or PCR amplification. Deletions in a particular *ZHMUP-2* exon can be detected using PCR primers that flank the exon. Table 1 provides the locations of *ZHMUP-2* exons present in the nucleotide sequences of SEQ ID NOs:1 and 5. This information can be used to design primers that amplify particular 25 exons.

30 Mutations can also be detected by hybridizing an oligonucleotide probe comprising a normal *ZHMUP-2* sequence to a Southern blot or to membrane-bound PCR products. Discrimination is achieved by hybridizing under conditions of high stringency, or by washing under varying conditions of stringency. This analysis can be targeted to a particular coding sequence. Alternatively, this approach is used to examine splice-donor or splice-acceptor sites in the immediate flanking intron sequences, where disease-causing mutations are often located. Suitable oligonucleotides can be designed by extending the sequence into an exon of choice, using the information of Table 1 and SEQ ID NOs:1 and 5

35 The duplication of all or part of a gene can cause a disorder when the insertion of the duplicated material is inserted into the reading frame of a gene and causes premature termination of translation. The effect of such duplication can be

detected with the protein truncation assay described below. Duplication and insertion can be examined directly by analyzing a subject's genomic DNA with standard methods, such as Southern hybridization, fluorescence *in situ* hybridization, pulsed-field gel analysis, or PCR.

5           A point mutation can lead to a nonconservative change resulting in the alteration of *ZHMUP-2* function or a change of an amino acid codon to a stop codon. If a point mutation occurs within an intron, the mutation may affect the fidelity of splicing. A point mutation can be detected using standard techniques, such as Southern hybridization analysis, PCR analysis, sequencing, ligation chain reaction, and other  
10           approaches. In single-strand conformation polymorphism analysis, for example, fragments amplified by PCR are separated into single strands and fractionated by polyacrylamide gel electrophoresis under denaturing conditions. The rate of migration through the gel is a function of conformation, which depends upon the base sequence. A mutation can alter the rate of migration of one or both single strands. In a chemical  
15           cleavage approach, hybrid molecules are produced between test and control DNA (*e.g.*, DNA that encodes the amino acid sequence of SEQ ID NO:2). Sites of base pair mismatch due to a mutation will be mispaired, and the strands will be susceptible to chemical cleavage at these sites.

          The protein truncation test is also useful for detecting the inactivation of a  
20           gene in which translation-terminating mutations produce only portions of the encoded protein (see, for example, Stoppa-Lyonnet *et al.*, *Blood* 91:3920 (1998)). According to this approach, RNA is isolated from a biological sample, and used to synthesize cDNA. PCR is then used to amplify the *ZHMUP-2* target sequence and to introduce an RNA polymerase promoter, a translation initiation sequence, and an in-frame ATG triplet.  
25           PCR products are transcribed using an RNA polymerase, and the transcripts are translated *in vitro* with a T7-coupled reticulocyte lysate system. The translation products are then fractionated by SDS-PAGE to determine the lengths of the translation products. The protein truncation test is described, for example, by Dracopoli *et al.* (eds.), *Current Protocols in Human Genetics*, pages 9.11.1 - 9.11.18 (John Wiley & Sons 1998).

30           In an alternative approach, a mutation can be detected using ribonuclease A, which will cleave the RNA strand of an RNA-DNA hybrid at the site of a sequence mismatch. Briefly, a PCR-amplified sequence of a *ZHMUP-2* gene or cDNA of a subject is hybridized with *in vitro* transcribed labeled RNA probes prepared from the DNA of a normal, healthy individual chosen from the general population. The RNA-  
35           DNA hybrids are digested with ribonuclease A and analyzed using denaturing gel electrophoresis. Sequence mismatches between the two strands will cause cleavage of the protected fragment, and small additional fragments will be detected in the samples



derived from a subject who has a mutated *ZHMUP-2* gene. The site of mutation can be deduced from the sizes of the cleavage products.

Analysis of chromosomal DNA using the *ZHMUP-2* polynucleotide sequence is useful for correlating disease with abnormalities localized to chromosome 6q, in particular to chromosome 6p11.2 – 21. In one embodiment, the methods of the present invention provide a method of detecting a chromosome 6p11.2 – 21 abnormality in a sample from an individual comprising: (a) obtaining *ZHMUP-2* RNA from the sample, (b) generating *ZHMUP-2* cDNA by polymerase chain reaction, and (c) comparing the nucleotide sequence of the *ZHMUP-2* cDNA to the nucleic acid sequence as shown in SEQ ID NO:1. In further embodiments, the difference between the sequence of the *ZHMUP-2*cDNA or *ZHMUP-2* gene in the sample and the *ZHMUP-2* sequence as shown in SEQ ID NOs:1 or 5 is indicative of chromosome 6p11.2 – 21 abnormality.

In another embodiment, the present invention provides methods for detecting in a sample from an individual, a chromosome 6p11.2 – 21 abnormality, comprising the steps of: (a) contacting nucleic acid molecules of the sample with a nucleic acid probe that hybridizes with a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, its complements or fragments, under stringent conditions, and (b) detecting the presence or absence of hybridization of the probe with nucleic acid molecules in the sample, wherein the absence of hybridization is indicative of a chromosome 6p11.2 – 21 abnormality.

The present invention also provides methods of detecting in a sample from an individual, a *ZHMUP-2* gene abnormality associated with a disease, comprising: (a) isolating nucleic acid molecules that encode *ZHMUP-2* from the sample, and (b) comparing the nucleotide sequence of the isolated *ZHMUP-2*-encoding sequence with the nucleotide sequence of SEQ ID NOs:1 or 5, wherein the difference between the sequence of the isolated *ZHMUP-2*-encoding sequence or a polynucleotide encoding the *ZHMUP-2* polypeptide generated from the isolated *ZHMUP-2*-encoding sequence and the nucleotide sequence of SEQ ID NOs:1 or 5 is indicative of an *ZHMUP-2* gene abnormality.

The present invention also provides methods of detecting in a sample from a individual, an abnormality in expression of the *ZHMUP-2* gene, comprising: (a) obtaining *ZHMUP-2* RNA from the sample, (b) generating *ZHMUP-2* cDNA by polymerase chain reaction from the *ZHMUP-2* RNA, and (c) comparing the nucleotide sequence of the *ZHMUP-2* cDNA to the nucleotide sequence of SEQ ID NO:1, wherein a difference between the sequence of the *ZHMUP-2* cDNA and the nucleotide sequence of SEQ ID NO:1 is indicative of an abnormality in expression of the *ZHMUP-2* gene.

In other aspects, the present invention provides methods for detecting in a sample from an individual, a *ZHMUP-2* gene abnormality associated with a disease, comprising: (a) contacting sample nucleic acid molecules with a nucleic acid probe, wherein the probe hybridizes to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, its complements or fragments, under stringent conditions, and (b) detecting the presence or absence of hybridization is indicative of a *ZHMUP-2* abnormality.

*In situ* hybridization provides another approach for identifying *ZHMUP-2* gene abnormalities. According to this approach, a *ZHMUP-2* probe is labeled with a detectable marker by any method known in the art. For example, the probe can be directly labeled by random priming, end labeling, PCR, or nick translation. Suitable direct labels include radioactive labels such as  $^{32}\text{P}$ ,  $^3\text{H}$ , and  $^{35}\text{S}$  and non-radioactive labels such as fluorescent markers (e.g., fluorescein, Texas Red, AMCA blue (7-amino-4-methyl-coumanine-3-acetate), lucifer yellow, rhodamine, etc.), cyanin dyes, which are detectable with visible light, enzymes, and the like. Probes labeled with an enzyme can be detected through a colorimetric reaction by providing a substrate for the enzyme. In the presence of various substrates, different colors are produced by the reaction, and these colors can be visualized to separately detect multiple probes if desired. Suitable substrates for alkaline phosphatase include 5-bromo-4-chloro-3-indolylphosphate and nitro blue tetrazolium. One suitable substrate for horseradish peroxidase is diaminobenzoate.

An illustrative method for detecting chromosomal abnormalities with *in situ* hybridization is described by Wang *et al.*, U.S. patent No. 5,856,089. Following this approach, for example, a method of performing *in situ* hybridization with a *ZHMUP-2* probe to detect a chromosome structural abnormality in a cell from a fixed tissue sample obtained from a subject can comprise the steps of: (1) obtaining a fixed tissue sample from the patient, (2) pretreating the fixed tissue sample obtained in step (1) with a bisulfite ion composition, (3) digesting the fixed tissue sample with proteinase, (4) performing *in situ* hybridization on cells obtained from the digested fixed tissue sample of step (3) with a probe which specifically hybridizes to the *ZHMUP-2* gene, wherein a signal pattern of hybridized probes is obtained, (5) comparing the signal pattern of the hybridized probe in step (4) to a predetermined signal pattern of the hybridized probe obtained when performing *in situ* hybridization on cells having a normal critical chromosome region of interest, and (6) detecting a chromosome structural abnormality in the patient's cells, by detecting a difference between the signal pattern obtained in step (4) and the predetermined signal pattern. Examples of *ZHMUP-2* gene abnormalities include deletions, amplifications, translocations, inversions, and the like.

Further localization of the *ZHMUP-2* gene can be achieved using radiation hybrid mapping, which is a somatic cell genetic technique developed for constructing high-resolution, contiguous maps of mammalian chromosomes (Cox *et al.*, *Science* 250:245 (1990)). Partial or full knowledge of a gene's sequence allows one to design PCR primers suitable for use with chromosomal radiation hybrid mapping panels. Radiation hybrid mapping panels are commercially available which cover the entire human genome, such as the Stanford G3 RH Panel and the GeneBridge 4 RH Panel (Research Genetics, Inc., Huntsville, AL). These panels enable rapid, PCR-based chromosomal localizations and ordering of genes, sequence-tagged sites, and other nonpolymorphic and polymorphic markers within a region of interest. This includes establishing directly proportional physical distances between newly discovered genes of interest and previously mapped markers.

The present invention contemplates kits for performing a diagnostic assay for *ZHMUP-2* gene expression or to analyze the *ZHMUP-2* locus of a subject. Such kits comprise nucleic acid probes, such as double-stranded nucleic acid molecules comprising the nucleotide sequence of SEQ ID NO:1, or a fragment thereof, as well as single-stranded nucleic acid molecules having the complement of the nucleotide sequence of SEQ ID NO:1, or a fragment thereof. An illustrative fragment resides within nucleotides 46 to 555 of SEQ ID NO:1. Probe molecules may be DNA, RNA, oligonucleotides, and the like. Kits may comprise nucleic acid primers for performing PCR.

Such a kit can contain all the necessary elements to perform a nucleic acid diagnostic assay described above. A kit will comprise at least one container comprising a *ZHMUP-2* probe or primer. The kit may also comprise a second container comprising one or more reagents capable of indicating the presence of *ZHMUP-2* sequences. Examples of such indicator reagents include detectable labels such as radioactive labels, fluorochromes, chemiluminescent agents, and the like. A kit may also comprise a means for conveying to the user that the *ZHMUP-2* probes and primers are used to detect *ZHMUP-2* gene expression. For example, written instructions may state that the enclosed nucleic acid molecules can be used to detect either a nucleic acid molecule that encodes *ZHMUP-2*, or a nucleic acid molecule having a nucleotide sequence that is complementary to a *ZHMUP-2*-encoding nucleotide sequence, or to analyze chromosomal sequences associated with the *ZHMUP-2* locus. The written material can be applied directly to a container, or the written material can be provided in the form of a packaging insert.

### 10. Use of Anti-ZHMUP-2 Antibodies to Detect ZHMUP-2 Protein

The present invention contemplates the use of anti-ZHMUP-2 antibodies to screen biological samples *in vitro* for the presence of ZHMUP-2. In one type of *in vitro* assay, anti-ZHMUP-2 antibodies are used in liquid phase. For example, the presence of ZHMUP-2 in a biological sample can be tested by mixing the biological sample with a trace amount of labeled ZHMUP-2 and an anti-ZHMUP-2 antibody under conditions that promote binding between ZHMUP-2 and its antibody. Complexes of ZHMUP-2 and anti-ZHMUP-2 in the sample can be separated from the reaction mixture by contacting the complex with an immobilized protein which binds with the antibody, such as an Fc antibody or *Staphylococcus* protein A. The concentration of ZHMUP-2 in the biological sample will be inversely proportional to the amount of labeled ZHMUP-2 bound to the antibody and directly related to the amount of free labeled ZHMUP-2.

Alternatively, *in vitro* assays can be performed in which anti-ZHMUP-2 antibody is bound to a solid-phase carrier. For example, antibody can be attached to a polymer, such as aminodextran, in order to link the antibody to an insoluble support such as a polymer-coated bead, a plate or a tube. Other suitable *in vitro* assays will be readily apparent to those of skill in the art.

In another approach, anti-ZHMUP-2 antibodies can be used to detect ZHMUP-2 in tissue sections prepared from a biopsy specimen. Such immunochemical detection can be used to determine the relative abundance of ZHMUP-2 and to determine the distribution of ZHMUP-2 in the examined tissue. General immunochemistry techniques are well established (see, for example, Ponder, "Cell Marking Techniques and Their Application," in *Mammalian Development: A Practical Approach*, Monk (ed.), pages 115-38 (IRL Press 1987), Coligan at pages 5.8.1-5.8.8, Ausubel (1995) at pages 14.6.1 to 14.6.13 (Wiley Interscience 1990), and Manson (ed.), *Methods In Molecular Biology, Vol.10: Immunochemical Protocols* (The Humana Press, Inc. 1992)).

Immunochemical detection can be performed by contacting a biological sample with an anti-ZHMUP-2 antibody, and then contacting the biological sample with a detectably labeled molecule which binds to the antibody. For example, the detectably labeled molecule can comprise an antibody moiety that binds to anti-ZHMUP-2 antibody. Alternatively, the anti-ZHMUP-2 antibody can be conjugated with avidin/streptavidin (or biotin) and the detectably labeled molecule can comprise biotin (or avidin/streptavidin). Numerous variations of this basic technique are well-known to those of skill in the art.

Alternatively, an anti-ZHMUP-2 antibody can be conjugated with a detectable label to form an anti-ZHMUP-2 immunoconjugate. Suitable detectable labels include, for example, a radioisotope, a fluorescent label, a chemiluminescent label, an enzyme label, a bioluminescent label or colloidal gold. Methods of making and detecting

such detectably-labeled immunoconjugates are well-known to those of ordinary skill in the art, and are described in more detail below.

The detectable label can be a radioisotope that is detected by autoradiography. Isotopes that are particularly useful for the purpose of the present invention are  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  and  $^{14}\text{C}$ .

Anti-ZHMUP-2 immunoconjugates can also be labeled with a fluorescent compound. The presence of a fluorescently-labeled antibody is determined by exposing the immunoconjugate to light of the proper wavelength and detecting the resultant fluorescence. Fluorescent labeling compounds include fluorescein isothiocyanate, rhodamine, phycoerytherin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

Alternatively, anti-ZHMUP-2 immunoconjugates can be detectably labeled by coupling an antibody component to a chemiluminescent compound. The presence of the chemiluminescent-tagged immunoconjugate is determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of chemiluminescent labeling compounds include luminol, isoluminol, an aromatic acridinium ester, an imidazole, an acridinium salt and an oxalate ester.

Similarly, a bioluminescent compound can be used to label anti-ZHMUP-2 immunoconjugates of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Bioluminescent compounds that are useful for labeling include luciferin, luciferase and aequorin.

Alternatively, anti-ZHMUP-2 immunoconjugates can be detectably labeled by linking an anti-ZHMUP-2 antibody component to an enzyme. When the anti-ZHMUP-2-enzyme conjugate is incubated in the presence of the appropriate substrate, the enzyme moiety reacts with the substrate to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorometric or visual means. Examples of enzymes that can be used to detectably label polyspecific immunoconjugates include  $\beta$ -galactosidase, glucose oxidase, peroxidase and alkaline phosphatase.

Those of skill in the art will know of other suitable labels which can be employed in accordance with the present invention. The binding of marker moieties to anti-ZHMUP-2 antibodies can be accomplished using standard techniques known to the art. Typical methodology in this regard is described by Kennedy *et al.*, *Clin. Chim. Acta* 70:1 (1976), Schurs *et al.*, *Clin. Chim. Acta* 81:1 (1977), Shih *et al.*, *Int'l J. Cancer* 46:1101 (1990), Stein *et al.*, *Cancer Res.* 50:1330 (1990), and Coligan, *supra*.

Moreover, the convenience and versatility of immunochemical detection can be enhanced by using anti-ZHMUP-2 antibodies that have been conjugated with

avidin, streptavidin, and biotin (see, for example, Wilchek *et al.* (eds.), "Avidin-Biotin Technology," *Methods In Enzymology*, Vol. 184 (Academic Press 1990), and Bayer *et al.*, "Immunochemical Applications of Avidin-Biotin Technology," in *Methods In Molecular Biology*, Vol. 10, Manson (ed.), pages 149-162 (The Humana Press, Inc. 1992).

5       Methods for performing immunoassays are well-established. See, for example, Cook and Self, "Monoclonal Antibodies in Diagnostic Immunoassays," in *Monoclonal Antibodies: Production, Engineering, and Clinical Application*, Ritter and Ladyman (eds.), pages 180-208, (Cambridge University Press, 1995), Perry, "The Role of Monoclonal Antibodies in the Advancement of Immunoassay Technology," in *Monoclonal Antibodies: Principles and Applications*, Birch and Lennox (eds.), pages 107-120 (Wiley-Liss, Inc. 1995), and Diamandis, *Immunoassay* (Academic Press, Inc. 1996).

10       In a related approach, biotin- or FITC-labeled ZHMUP-2 can be used to identify cells that bind ZHMUP-2. Such can binding can be detected, for example, using flow cytometry.

15       The present invention contemplates kits for performing an immunological diagnostic assay for ZHMUP-2 gene expression. Such kits comprise at least one container comprising an anti-ZHMUP-2 antibody, or antibody fragment. A kit may also comprise a second container comprising one or more reagents capable of indicating the presence of ZHMUP-2 antibody or antibody fragments. Examples of such indicator reagents include  
20       detectable labels such as a radioactive label, a fluorescent label, a chemiluminescent label, an enzyme label, a bioluminescent label, colloidal gold, and the like. A kit may also comprise a means for conveying to the user that ZHMUP-2 antibodies or antibody fragments are used to detect ZHMUP-2 protein. For example, written instructions may state that the enclosed antibody or antibody fragment can be used to detect ZHMUP-2.  
25       The written material can be applied directly to a container, or the written material can be provided in the form of a packaging insert.

30       In addition to the detection kits described above, polynucleotides and polypeptides of the present invention will be useful as educational tools in laboratory practicum kits for courses related to genetics and molecular biology, protein chemistry, and antibody production and analysis. Due to its unique polynucleotide and polypeptide sequences, molecules of ZHMUP-2 can be used as standards or as "unknowns" for testing purposes. For example, ZHMUP-2 polynucleotides can be used as an aid, such as, for example, to teach a student how to prepare expression constructs for bacterial, viral, or mammalian expression, including fusion constructs, wherein ZHMUP-2 is the  
35       gene to be expressed; for determining the restriction endonuclease cleavage sites of the polynucleotides; determining mRNA and DNA localization of ZHMUP-2 polynucleotides in tissues (*i.e.*, by northern and Southern blotting as well as polymerase

chain reaction); and for identifying related polynucleotides and polypeptides by nucleic acid hybridization. As an illustration, students will find that *Pst*I digestion of a nucleic acid molecule consisting of the nucleotide sequence of SEQ ID NO:1 provides fragments of 140 base pairs, and 415 base pairs, and that digestion with *Eco*RI yields fragments of 202 base pairs, and 353 base pairs.

ZHMUP-2 polypeptides can be used as an aid to teach preparation of antibodies; identifying proteins by western blotting; protein purification; determining the weight of produced ZHMUP-2 polypeptides as a ratio to total protein produced; identifying peptide cleavage sites; coupling amino and carboxyl terminal tags; amino acid sequence analysis, as well as, but not limited to monitoring biological activities of both the native and tagged protein (*i.e.*, protease inhibition) *in vitro* and *in vivo*. For example, students will find that digestion of an unglycosylated ZHMUP-2 polypeptide consisting of the amino acid sequence of SEQ ID NO:2 with BNPS or NCS/urea yields fragments having approximate molecular weights of 4101, and 17099, whereas digestion of such a polypeptide with hydroxylamine provides six fragments with approximate molecular weights of 9070, 749, 1002, 885, 3054, and 6507.

ZHMUP-2 polypeptides can also be used to teach analytical skills such as mass spectrometry, circular dichroism to determine conformation, especially of the four alpha helices, x-ray crystallography to determine the three-dimensional structure in atomic detail, nuclear magnetic resonance spectroscopy to reveal the structure of proteins in solution. For example, a kit containing the ZHMUP-2 can be given to the student to analyze. Since the amino acid sequence would be known by the instructor, the protein can be given to the student as a test to determine the skills or develop the skills of the student, the instructor would then know whether or not the student has correctly analyzed the polypeptide. Since every polypeptide is unique, the educational utility of ZHMUP-2 would be unique unto itself.

The antibodies which bind specifically to ZHMUP-2 can be used as a teaching aid to instruct students how to prepare affinity chromatography columns to purify ZHMUP-2, cloning and sequencing the polynucleotide that encodes an antibody and thus as a practicum for teaching a student how to design humanized antibodies. The ZHMUP-2 gene, polypeptide, or antibody would then be packaged by reagent companies and sold to educational institutions so that the students gain skill in art of molecular biology. Because each gene and protein is unique, each gene and protein creates unique challenges and learning experiences for students in a lab practicum. Such educational kits containing the ZHMUP-2 gene, polypeptide, or antibody are considered within the scope of the present invention.

## 11. ZHMUP-2 Analogs, Receptors, and Ligands

One general class of ZHMUP-2 analogs are ZHMUP-2 variants having an amino acid sequence that is a mutation of the amino acid sequence disclosed herein. Another general class of ZHMUP-2 analogs is provided by anti-idiotypic antibodies, and fragments thereof. Moreover, recombinant antibodies comprising anti-idiotypic variable domains can be used as analogs (see, for example, Monfardini *et al.*, *Proc. Assoc. Am. Physicians* 108:420 (1996)). Since the variable domains of anti-idiotypic ZHMUP-2 antibodies mimic ZHMUP-2, these domains can provide either ZHMUP-2 agonist or antagonist activity. As an illustration, Lim and Langer, *J. Interferon Res.* 13:295 (1993), describe anti-idiotypic interferon- $\alpha$  antibodies that have the properties of either interferon- $\alpha$  agonists or antagonists.

Another approach to identifying ZHMUP-2 analogs is provided by the use of combinatorial libraries. Methods for constructing and screening phage display and other combinatorial libraries are provided, for example, by Kay *et al.*, *Phage Display of Peptides and Proteins* (Academic Press 1996), Verdine, U.S. Patent No. 5,783,384, Kay, *et al.*, U.S. Patent No. 5,747,334, and Kauffman *et al.*, U.S. Patent No. 5,723,323.

The ZHMUP-2 polypeptides of the present invention can be used to identify small molecules that bind ZHMUP-2 ("a ZHMUP-2 ligand"), as well as proteins that bind with ZHMUP-2 ("a ZHMUP-2 receptor"). For example, ZHMUP-2 ligands can be identified by determining whether potential ligands bind with ZHMUP-2 polypeptides *in vitro*. Potential ZHMUP-2 ligands include members of the 16-androstenes, estrenes, and other putative human pheromones. In these assays, either the ZHMUP-2 ligand or the ZHMUP-2 polypeptide may be detectably labeled. General methods for performing binding assays are described above.

The location of ZHMUP-2 receptor binding domains can be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos *et al.*, *Science* 255:306 (1992), Smith *et al.*, *J. Mol. Biol.* 224:899 (1992), and Wlodaver *et al.*, *FEBS Lett.* 309:59 (1992).

Anti-idiotypic ZHMUP-2 antibodies, as well as ZHMUP-2 polypeptides can be used to identify and to isolate ZHMUP-2 receptors. For example, proteins and peptides of the present invention can be immobilized on a column and used to bind receptor proteins from membrane preparations that are run over the column (Hermanson *et al.* (eds.), *Immobilized Affinity Ligand Techniques*, pages 195-202 (Academic Press 1992)). Also see, Varthakavi and Minocha, *J. Gen. Virol.* 77:1875 (1996), who describe the use of anti-idiotypic antibodies for receptor identification. In another approach,



receptor proteins that bind ZHMUP-2 can be isolated from cell membranes by photocrosslinking, solubilizing, and then immunoprecipitating ZHMUP-2 and ZHMUP-2 receptor complexes using antibodies to ZHMUP-2.

5 Radiolabeled or affinity labeled ZHMUP-2 polypeptides can also be used to identify or to localize ZHMUP-2 receptors in a biological sample (see, for example, Deutscher (ed.), *Methods in Enzymol.*, vol. 182, pages 721-37 (Academic Press 1990); Brunner *et al.*, *Ann. Rev. Biochem.* 62:483 (1993); Fedan *et al.*, *Biochem. Pharmacol.* 33:1167 (1984)). Moreover, ZHMUP-2 labeled with biotin or FITC can be used for expression cloning of ZHMUP-2 receptors. Alternatively, a cDNA encoding a ZHMUP-2 receptor can be isolated from a vomeronasal organ cDNA library by expression cloning  
10 protocols similar to those described by Jelinek *et al.*, *Science* 259:1614 (1993).

Those of skill in the art can devise various methods to measure the ability of ZHMUP-2 polypeptides, with or without a ZHMUP-2 ligand, to induce physiological effects. For example, human postmortum vomeronasal membranes for signal  
15 transduction studies can be isolated employing a method described for rodent vomeronasal membrane preparations (Kroner *et al.*, *Neuroport* 7:2989 (1996)). Moreover, stimulation experiments and second messenger assays, performed with recombinant ZHMUP-2 alone or in combination with ligand, can be carried out employing the method described by Krieger *et al.*, *J. Biol. Chem.* 274:4655 (1999).  
20 Formulations of ZHMUP-2 alone or in combination with ligand, can also be assayed on vomeronasal organs of human volunteers as described by Monti-Bloch and Grosser, *J. Steroid Biochem.* 39:573 (1991), and by Grosser *et al.*, *Psychoneuroendocrinology* 25:289 (2000). These assays can be used to assess changes in the electrophysiological output of the vomeronasal organ, as well as alternations in autonomic functions, and  
25 changes in transient feelings and moods. Alternations of hypothalamic functions, such as satiety, energy balance, sexual motivation, anxiety and the like, can also be evaluated in test subjects using a variety of recognized standard test protocols. Useful formulations of ZHMUP-2 can be conveniently delivered to vomeronasal organ by intranasal administration.

30 In another approach, a ZHMUP-2 polypeptide or ZHMUP-2 fusion protein can be immobilized onto the surface of a receptor chip of a biosensor instrument (BIACORE, Biacore AB; Uppsala, Sweden) to detect the presence of a ZHMUP-2 target, such as a ZHMUP-2 receptor or a ZHMUP-2 ligand. The use of this instrument is disclosed, for example, by Karlsson, *Immunol. Methods* 145:229 (1991). In brief, a  
35 ZHMUP-2 polypeptide or fusion protein is covalently attached, using amine or sulfhydryl chemistry, to dextran fibers that are attached to gold film within a flow cell. A test sample is then passed through the cell. If a ZHMUP-2 target molecule is present

in the sample, it will bind to the immobilized polypeptide or fusion protein, causing a change in the refractive index of the medium, which is detected as a change in surface plasmon resonance of the gold film. This system allows the determination on- and off-rates, from which binding affinity can be calculated, and assessment of the stoichiometry of binding, as well as the kinetic effects of ZHMUP-2 mutation.

## 12. Therapeutic Uses of Polypeptides Having ZHMUP-2 Activity

The present invention includes the use of proteins, polypeptides, and peptides having ZHMUP-2 activity (such as ZHMUP-2 polypeptides, anti-idiotypic anti-ZHMUP-2 antibodies, and ZHMUP-2 fusion proteins) to a subject who lacks an adequate amount of this polypeptide. The ZHMUP-2 molecules described herein can be administered, with or without a cognate pheromone ligand, to any subject in need of treatment, and the present invention contemplates both veterinary and human therapeutic uses. Illustrative subjects include mammalian subjects, such as farm animals, domestic animals, and human patients.

Generally, the dosage of administered polypeptide, protein or peptide will vary depending upon such factors as the subject's age, weight, height, sex, general medical condition and previous medical history. Typically, it is desirable to provide the recipient with a dosage of a molecule having ZHMUP-2 activity, which is in the range of from about 1 pg/kg to 10 mg/kg (amount of agent/body weight of subject), although a lower or higher dosage also may be administered as circumstances dictate.

Molecules having ZHMUP-2 activity can be administered to a subject by oral, dermal, mucosal-membrane, pulmonary, and transcutaneous routes. Oral delivery is suitable for polyester microspheres, zein microspheres, proteinoid microspheres, polycyanoacrylate microspheres, and lipid-based systems (see, for example, DiBase and Morrel, "Oral Delivery of Microencapsulated Proteins," in *Protein Delivery: Physical Systems*, Sanders and Hendren (eds.), pages 255-288 (Plenum Press 1997)). The feasibility of an intranasal delivery is exemplified by such a mode of insulin administration (see, for example, Hinchcliffe and Illum, *Adv. Drug Deliv. Rev.* 35:199 (1999)). Dry or liquid particles comprising ZHMUP-2 can be prepared and inhaled with the aid of dry-powder dispersers, liquid aerosol generators, or nebulizers (e.g., Pettit and Gombotz, *TIBTECH* 16:343 (1998); Patton *et al.*, *Adv. Drug Deliv. Rev.* 35:235 (1999)). This approach is illustrated by the AERX diabetes management system, which is a hand-held electronic inhaler that delivers aerosolized insulin into the lungs. Studies have shown that proteins as large as 48,000 kDa have been delivered across skin at therapeutic concentrations with the aid of low-frequency ultrasound, which illustrates the feasibility

of transcutaneous administration (Mitragotri *et al.*, *Science* 269:850 (1995)). Transdermal delivery using electroporation provides another means to administer ZHMUP-2 (Potts *et al.*, *Pharm. Biotechnol.* 10:213 (1997)).

5 A molecule having ZHMUP-2 activity can also be administered to a subject by intravenous, intraarterial, intraperitoneal, intramuscular, subcutaneous, intrapleural, or intrathecal routes, or by perfusion through a regional catheter. When administering therapeutic proteins by injection, the administration may be by continuous infusion or by single or multiple boluses.

10 A pharmaceutical composition comprising a protein, polypeptide, or peptide having ZHMUP-2 activity can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby the therapeutic proteins are combined in a mixture with a pharmaceutically acceptable carrier. A composition is said to be a "pharmaceutically acceptable carrier" if its administration can be tolerated by a recipient subject. Sterile phosphate-buffered saline is one example of a pharmaceutically  
15 acceptable carrier. Other suitable carriers are well-known to those in the art. See, for example, Gennaro (ed.), *Remington's Pharmaceutical Sciences*, 19th Edition (Mack Publishing Company 1995).

For purposes of therapy, molecules having ZHMUP-2 activity and a pharmaceutically acceptable carrier are administered to a subject in a therapeutically  
20 effective amount. A combination of a protein, polypeptide, or peptide having ZHMUP-2 activity and a pharmaceutically acceptable carrier is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient subject.

25 A pharmaceutical composition comprising molecules having ZHMUP-2 activity can be furnished in liquid form, or in solid form. Liquid forms, including liposome-encapsulated formulations, are illustrated by injectable solutions and oral suspensions. Exemplary solid forms include capsules, tablets, and controlled-release forms, such as a miniosmotic pump or an implant. Other dosage forms can be devised  
30 by those skilled in the art, as shown, for example, by Ansel and Popovich, *Pharmaceutical Dosage Forms and Drug Delivery Systems*, 5<sup>th</sup> Edition (Lea & Febiger 1990), Gennaro (ed.), *Remington's Pharmaceutical Sciences*, 19<sup>th</sup> Edition (Mack Publishing Company 1995), and by Ranade and Hollinger, *Drug Delivery Systems* (CRC Press 1996).

35 As an illustration, ZHMUP-2 pharmaceutical compositions may be supplied as a kit comprising a container that comprises ZHMUP-2. ZHMUP-2 can be provided in the form of an injectable solution for single or multiple doses, or as a sterile

powder that will be reconstituted before injection. Such a kit may further comprise written information on indications and usage of the pharmaceutical composition. Moreover, such information may include a statement that the ZHMUP-2 composition is contraindicated in subjects with known hypersensitivity to ZHMUP-2.

5

### 13. Therapeutic Uses of ZHMUP-2 Nucleotide Sequences

The present invention includes the use of ZHMUP-2 nucleotide sequences to provide ZHMUP-2 to a subject in need of such treatment. In addition, a therapeutic expression vector can be provided that inhibits *ZHMUP-2* gene expression, such as an anti-sense molecule, a ribozyme, or an external guide sequence molecule.

There are numerous approaches to introduce a *ZHMUP-2* gene to a subject, including the use of recombinant host cells that express ZHMUP-2, delivery of naked nucleic acid encoding ZHMUP-2, use of a cationic lipid carrier with a nucleic acid molecule that encodes ZHMUP-2, and the use of viruses that express ZHMUP-2, such as recombinant retroviruses, recombinant adeno-associated viruses, recombinant adenoviruses, and recombinant Herpes simplex viruses [HSV] (see, for example, Mulligan, *Science* 260:926 (1993), Rosenberg *et al.*, *Science* 242:1575 (1988), LaSalle *et al.*, *Science* 259:988 (1993), Wolff *et al.*, *Science* 247:1465 (1990), Breakfield and Deluca, *The New Biologist* 3:203 (1991)). In an *ex vivo* approach, for example, cells are isolated from a subject, transfected with a vector that expresses a *ZHMUP-2* gene, and then transplanted into the subject.

In order to effect expression of a *ZHMUP-2* gene, an expression vector is constructed in which a nucleotide sequence encoding a *ZHMUP-2* gene is operably linked to a core promoter, and optionally a regulatory element, to control gene transcription. The general requirements of an expression vector are described above.

Alternatively, a *ZHMUP-2* gene can be delivered using recombinant viral vectors, including for example, adenoviral vectors (*e.g.*, Kass-Eisler *et al.*, *Proc. Nat'l Acad. Sci. USA* 90:11498 (1993), Kolls *et al.*, *Proc. Nat'l Acad. Sci. USA* 91:215 (1994), Li *et al.*, *Hum. Gene Ther.* 4:403 (1993), Vincent *et al.*, *Nat. Genet.* 5:130 (1993), and Zabner *et al.*, *Cell* 75:207 (1993)), adenovirus-associated viral vectors (Flotte *et al.*, *Proc. Nat'l Acad. Sci. USA* 90:10613 (1993)), alphaviruses such as Semliki Forest Virus and Sindbis Virus (Hertz and Huang, *J. Vir.* 66:857 (1992), Raju and Huang, *J. Vir.* 65:2501 (1991), and Xiong *et al.*, *Science* 243:1188 (1989)), herpes viral vectors (*e.g.*, U.S. Patent Nos. 4,769,331, 4,859,587, 5,288,641 and 5,328,688), parvovirus vectors (Koering *et al.*, *Hum. Gene Therap.* 5:457 (1994)), pox virus vectors (Ozaki *et al.*, *Biochem. Biophys. Res. Comm.* 193:653 (1993), Panicali and Paoletti, *Proc. Nat'l Acad.*

5 *Sci. USA* 79:4927 (1982)), pox viruses, such as canary pox virus or vaccinia virus (Fisher-Hoch *et al.*, *Proc. Nat'l Acad. Sci. USA* 86:317 (1989), and Flexner *et al.*, *Ann. N.Y. Acad. Sci.* 569:86 (1989)), and retroviruses (*e.g.*, Baba *et al.*, *J. Neurosurg* 79:729 (1993), Ram *et al.*, *Cancer Res.* 53:83 (1993), Takamiya *et al.*, *J. Neurosci. Res* 33:493 (1992), Vile and Hart, *Cancer Res.* 53:962 (1993), Vile and Hart, *Cancer Res.* 53:3860 (1993), and Anderson *et al.*, U.S. Patent No. 5,399,346). Within various embodiments, either the viral vector itself, or a viral particle which contains the viral vector may be utilized in the methods and compositions described below.

10 As an illustration of one system, adenovirus, a double-stranded DNA virus, is a well-characterized gene transfer vector for delivery of a heterologous nucleic acid molecule (for a review, see Becker *et al.*, *Meth. Cell Biol.* 43:161 (1994); Douglas and Curiel, *Science & Medicine* 4:44 (1997)). The adenovirus system offers several advantages including: (i) the ability to accommodate relatively large DNA inserts, (ii) the ability to be grown to high-titer, (iii) the ability to infect a broad range of mammalian cell types, and (iv) the ability to be used with many different promoters including ubiquitous, tissue specific, and regulatable promoters. In addition, adenoviruses can be administered by intravenous injection, because the viruses are stable in the bloodstream.

15 Using adenovirus vectors where portions of the adenovirus genome are deleted, inserts are incorporated into the viral DNA by direct ligation or by homologous recombination with a co-transfected plasmid. In an exemplary system, the essential E1 gene is deleted from the viral vector, and the virus will not replicate unless the E1 gene is provided by the host cell. When intravenously administered to intact animals, adenovirus primarily targets the liver. Although an adenoviral delivery system with an E1 gene deletion cannot replicate in the host cells, the host's tissue will express and process an encoded heterologous protein. Host cells will also secrete the heterologous protein if the corresponding gene includes a secretory signal sequence. Secreted proteins will enter the circulation from tissue that expresses the heterologous gene (*e.g.*, the highly vascularized liver).

20 Moreover, adenoviral vectors containing various deletions of viral genes can be used to reduce or eliminate immune responses to the vector. Such adenoviruses are E1-deleted, and in addition, contain deletions of E2A or E4 (Lusky *et al.*, *J. Virol.* 72:2022 (1998); Raper *et al.*, *Human Gene Therapy* 9:671 (1998)). The deletion of E2b has also been reported to reduce immune responses (Amalfitano *et al.*, *J. Virol.* 72:926 (1998)). By deleting the entire adenovirus genome, very large inserts of heterologous DNA can be accommodated. Generation of so called "gutless" adenoviruses, where all viral genes are deleted, are particularly advantageous for insertion of large inserts of heterologous DNA (for a review, see Yeh. and Perricaudet, *FASEB J.* 11:615 (1997)).

High titer stocks of recombinant viruses capable of expressing a therapeutic gene can be obtained from infected mammalian cells using standard methods. For example, recombinant HSV can be prepared in Vero cells, as described by Brandt *et al.*, *J. Gen. Virol.* 72:2043 (1991), Herold *et al.*, *J. Gen. Virol.* 75:1211 (1994), Visalli and Brandt, *Virology* 185:419 (1991), Grau *et al.*, *Invest. Ophthalmol. Vis. Sci.* 30:2474 (1989), Brandt *et al.*, *J. Virol. Meth.* 36:209 (1992), and by Brown and MacLean (eds.), *HSV Virus Protocols* (Humana Press 1997).

Alternatively, an expression vector comprising a ZHMUP-2 gene can be introduced into a subject's cells by lipofection *in vivo* using liposomes. Synthetic cationic lipids can be used to prepare liposomes for *in vivo* transfection of a gene encoding a marker (Felgner *et al.*, *Proc. Nat'l Acad. Sci. USA* 84:7413 (1987); Mackey *et al.*, *Proc. Nat'l Acad. Sci. USA* 85:8027 (1988)). The use of lipofection to introduce exogenous genes into specific organs *in vivo* has certain practical advantages. Liposomes can be used to direct transfection to particular cell types, which is particularly advantageous in a tissue with cellular heterogeneity, such as the pancreas, liver, kidney, and brain. Lipids may be chemically coupled to other molecules for the purpose of targeting. Targeted peptides (*e.g.*, hormones or neurotransmitters), proteins such as antibodies, or non-peptide molecules can be coupled to liposomes chemically.

Electroporation is another alternative mode of administration of a ZHMUP-2 nucleic acid molecules. For example, Aihara and Miyazaki, *Nature Biotechnology* 16:867 (1998), have demonstrated the use of *in vivo* electroporation for gene transfer into muscle.

In an alternative approach to gene therapy, a therapeutic gene may encode a ZHMUP-2 anti-sense RNA that inhibits the expression of ZHMUP-2. Methods of preparing anti-sense constructs are known to those in the art. See, for example, Erickson *et al.*, *Dev. Genet.* 14:274 (1993) [transgenic mice], Augustine *et al.*, *Dev. Genet.* 14:500 (1993) [murine whole embryo culture], and Olson and Gibo, *Exp. Cell Res.* 241:134 (1998) [cultured cells]. Suitable sequences for ZHMUP-2 anti-sense molecules can be derived from the nucleotide sequences of ZHMUP-2 disclosed herein.

Alternatively, an expression vector can be constructed in which a regulatory element is operably linked to a nucleotide sequence that encodes a ribozyme. Ribozymes can be designed to express endonuclease activity that is directed to a certain target sequence in a mRNA molecule (see, for example, Draper and Macejak, U.S. Patent No. 5,496,698, McSwiggen, U.S. Patent No. 5,525,468, Chowrira and McSwiggen, U.S. Patent No. 5,631,359, and Robertson and Goldberg, U.S. Patent No. 5,225,337). In the context of the present invention, ribozymes include nucleotide sequences that bind with ZHMUP-2 mRNA.

In another approach, expression vectors can be constructed in which a regulatory element directs the production of RNA transcripts capable of promoting RNase P-mediated cleavage of mRNA molecules that encode a *ZHMUP-2* gene. According to this approach, an external guide sequence can be constructed for directing the endogenous ribozyme, RNase P, to a particular species of intracellular mRNA, which is subsequently cleaved by the cellular ribozyme (see, for example, Altman *et al.*, U.S. Patent No. 5,168,053, Yuan *et al.*, *Science* 263:1269 (1994), Pace *et al.*, international publication No. WO 96/18733, George *et al.*, international publication No. WO 96/21731, and Werner *et al.*, international publication No. WO 97/33991). Preferably, the external guide sequence comprises a ten to fifteen nucleotide sequence complementary to *ZHMUP-2* mRNA, and a 3'-NCCA nucleotide sequence, wherein N is preferably a purine. The external guide sequence transcripts bind to the targeted mRNA species by the formation of base pairs between the mRNA and the complementary external guide sequences, thus promoting cleavage of mRNA by RNase P at the nucleotide located at the 5'-side of the base-paired region.

In general, the dosage of a composition comprising a therapeutic vector having a *ZHMUP-2* nucleotide acid sequence, such as a recombinant virus, will vary depending upon such factors as the subject's age, weight, height, sex, general medical condition and previous medical history. Suitable routes of administration of therapeutic vectors include intravenous injection, intraarterial injection, intraperitoneal injection, intramuscular injection, intratumoral injection, and injection into a cavity that contains a tumor.

A composition comprising viral vectors, non-viral vectors, or a combination of viral and non-viral vectors of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby vectors or viruses are combined in a mixture with a pharmaceutically acceptable carrier. As noted above, a composition, such as phosphate-buffered saline is said to be a "pharmaceutically acceptable carrier" if its administration can be tolerated by a recipient subject. Other suitable carriers are well-known to those in the art (see, for example, *Remington's Pharmaceutical Sciences, 19th Ed.* (Mack Publishing Co. 1995), and *Gilman's the Pharmacological Basis of Therapeutics, 7th Ed.* (MacMillan Publishing Co. 1985)).

For purposes of therapy, a therapeutic gene expression vector, or a recombinant virus comprising such a vector, and a pharmaceutically acceptable carrier are administered to a subject in a therapeutically effective amount. A combination of an expression vector (or virus) and a pharmaceutically acceptable carrier is said to be administered in a "therapeutically effective amount" if the amount administered is

physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient subject.

When the subject treated with a therapeutic gene expression vector or a recombinant virus is a human, then the therapy is preferably somatic cell gene therapy. That is, the preferred treatment of a human with a therapeutic gene expression vector or a recombinant virus does not entail introducing into cells a nucleic acid molecule that can form part of a human germ line and be passed onto successive generations (*i.e.*, human germ line gene therapy).

#### 14. Production of Transgenic Mice

Transgenic mice can be engineered to over-express the *ZHMUP-2* gene in all tissues or under the control of a tissue-specific or tissue-preferred regulatory element. These over-producers of *ZHMUP-2* can be used to characterize the phenotype that results from over-expression, and the transgenic animals can serve as models for human disease caused by excess *ZHMUP-2*. Transgenic mice that over-express *ZHMUP-2* also provide model bioreactors for production of *ZHMUP-2* in the milk or blood of larger animals. Methods for producing transgenic mice are well-known to those of skill in the art (see, for example, Jacob, "Expression and Knockout of Interferons in Transgenic Mice," in *Overexpression and Knockout of Cytokines in Transgenic Mice*, Jacob (ed.), pages 111-124 (Academic Press, Ltd. 1994), Monastersky and Robl (eds.), *Strategies in Transgenic Animal Science* (ASM Press 1995), and Abbud and Nilson, "Recombinant Protein Expression in Transgenic Mice," in *Gene Expression Systems: Using Nature for the Art of Expression*, Fernandez and Hoeffler (eds.), pages 367-397 (Academic Press, Inc. 1999)).

For example, a method for producing a transgenic mouse that expresses a *ZHMUP-2* gene can begin with adult, fertile males (studs) (B6C3f1, 2-8 months of age (Taconic Farms, Germantown, NY)), vasectomized males (duds) (B6D2f1, 2-8 months, (Taconic Farms)), prepubescent fertile females (donors) (B6C3f1, 4-5 weeks, (Taconic Farms)) and adult fertile females (recipients) (B6D2f1, 2-4 months, (Taconic Farms)). The donors are acclimated for one week and then injected with approximately 8 IU/mouse of Pregnant Mare's Serum gonadotrophin (Sigma Chemical Company; St. Louis, MO) I.P., and 46-47 hours later, 8 IU/mouse of human Chorionic Gonadotropin (hCG (Sigma)) I.P. to induce superovulation. Donors are mated with studs subsequent to hormone injections. Ovulation generally occurs within 13 hours of hCG injection. Copulation is confirmed by the presence of a vaginal plug the morning following mating. Fertilized eggs are collected under a surgical scope. The oviducts are collected and eggs



are released into urinalysis slides containing hyaluronidase (Sigma). Eggs are washed once in hyaluronidase, and twice in Whitten's W640 medium (described, for example, by Menino and O'Claray, *Biol. Reprod.* 77:159 (1986), and Dienhart and Downs, *Zygote* 4:129 (1996)) that has been incubated with 5% CO<sub>2</sub>, 5% O<sub>2</sub>, and 90% N<sub>2</sub> at 37°C. The  
5 eggs are then stored in a 37°C/5% CO<sub>2</sub> incubator until microinjection.

Ten to twenty micrograms of plasmid DNA containing a *ZHMUP-2* encoding sequence is linearized, gel-purified, and resuspended in 10 mM Tris-HCl (pH 7.4), 0.25 mM EDTA (pH 8.0), at a final concentration of 5-10 nanograms per microliter for microinjection. For example, the *ZHMUP-2* encoding sequences can encode a  
10 polypeptide comprising amino acid residues 16 to 185 of SEQ ID NO:2.

Plasmid DNA is microinjected into harvested eggs contained in a drop of W640 medium overlaid by warm, CO<sub>2</sub>-equilibrated mineral oil. The DNA is drawn into an injection needle (pulled from a 0.75mm ID, 1mm OD borosilicate glass capillary), and injected into individual eggs. Each egg is penetrated with the injection needle, into  
15 one or both of the haploid pronuclei.

Picoliters of DNA are injected into the pronuclei, and the injection needle withdrawn without coming into contact with the nucleoli. The procedure is repeated until all the eggs are injected. Successfully microinjected eggs are transferred into an organ tissue-culture dish with pre-gassed W640 medium for storage overnight in a  
20 37°C/5% CO<sub>2</sub> incubator.

The following day, two-cell embryos are transferred into pseudopregnant recipients. The recipients are identified by the presence of copulation plugs, after copulating with vasectomized duds. Recipients are anesthetized and shaved on the dorsal left side and transferred to a surgical microscope. A small incision is made in the  
25 skin and through the muscle wall in the middle of the abdominal area outlined by the ribcage, the saddle, and the hind leg, midway between knee and spleen. The reproductive organs are exteriorized onto a small surgical drape. The fat pad is stretched out over the surgical drape, and a baby serrefine (Roboz, Rockville, MD) is attached to the fat pad and left hanging over the back of the mouse, preventing the organs from sliding back in.

30 With a fine transfer pipette containing mineral oil followed by alternating W640 and air bubbles, 12-17 healthy two-cell embryos from the previous day's injection are transferred into the recipient. The swollen ampulla is located and holding the oviduct between the ampulla and the bursa, a nick in the oviduct is made with a 28 g needle close to the bursa, making sure not to tear the ampulla or the bursa.

35 The pipette is transferred into the nick in the oviduct, and the embryos are blown in, allowing the first air bubble to escape the pipette. The fat pad is gently pushed into the peritoneum, and the reproductive organs allowed to slide in. The peritoneal wall

is closed with one suture and the skin closed with a wound clip. The mice recuperate on a 37°C slide warmer for a minimum of four hours.

The recipients are returned to cages in pairs, and allowed 19-21 days gestation. After birth, 19-21 days postpartum is allowed before weaning. The weanlings are sexed and placed into separate sex cages, and a 0.5 cm biopsy (used for genotyping) is snipped off the tail with clean scissors.

Genomic DNA is prepared from the tail snips using, for example, a QIAGEN DNEASY kit following the manufacturer's instructions. Genomic DNA is analyzed by PCR using primers designed to amplify a *ZHMUP-2* gene or a selectable marker gene that was introduced in the same plasmid. After animals are confirmed to be transgenic, they are back-crossed into an inbred strain by placing a transgenic female with a wild-type male, or a transgenic male with one or two wild-type female(s). As pups are born and weaned, the sexes are separated, and their tails snipped for genotyping.

To check for expression of a transgene in a live animal, a partial hepatectomy is performed. A surgical prep is made of the upper abdomen directly below the zyphoid process. Using sterile technique, a small 1.5-2 cm incision is made below the sternum and the left lateral lobe of the liver exteriorized. Using 4-0 silk, a tie is made around the lower lobe securing it outside the body cavity. An atraumatic clamp is used to hold the tie while a second loop of absorbable Dexon (American Cyanamid; Wayne, N.J.) is placed proximal to the first tie. A distal cut is made from the Dexon tie and approximately 100 mg of the excised liver tissue is placed in a sterile petri dish. The excised liver section is transferred to a 14 ml polypropylene round bottom tube and snap frozen in liquid nitrogen and then stored on dry ice. The surgical site is closed with suture and wound clips, and the animal's cage placed on a 37°C heating pad for 24 hours post operatively. The animal is checked daily post operatively and the wound clips removed 7 to 10 days after surgery. The expression level of *ZHMUP-2* mRNA is examined for each transgenic mouse using an RNA solution hybridization assay or polymerase chain reaction.

In addition to producing transgenic mice that over-express *ZHMUP-2*, it is useful to engineer transgenic mice with either abnormally low or no expression of the gene. Such transgenic mice provide useful models for diseases associated with a lack of *ZHMUP-2*. As discussed above, *ZHMUP-2* gene expression can be inhibited using anti-sense genes, ribozyme genes, or external guide sequence genes. To produce transgenic mice that under-express the *ZHMUP-2* gene, such inhibitory sequences are targeted to *ZHMUP-2* mRNA. Methods for producing transgenic mice that have abnormally low expression of a particular gene are known to those in the art (see, for example, Wu *et al.*,

"Gene Underexpression in Cultured Cells and Animals by Antisense DNA and RNA Strategies," in *Methods in Gene Biotechnology*, pages 205-224 (CRC Press 1997)).

5 An alternative approach to producing transgenic mice that have little or no *ZHMUP-2* gene expression is to generate mice having at least one normal *ZHMUP-2* allele replaced by a nonfunctional *ZHMUP-2* gene. One method of designing a nonfunctional *ZHMUP-2* gene is to insert another gene, such as a selectable marker gene, within a nucleic acid molecule that encodes *ZHMUP-2*. Standard methods for producing these so-called "knockout mice" are known to those skilled in the art (see, for example, Jacob, "Expression and Knockout of Interferons in Transgenic Mice," in *Overexpression and Knockout of Cytokines in Transgenic Mice*, Jacob (ed.), pages 111-124 (Academic Press, Ltd. 1994), and Wu *et al.*, "New Strategies for Gene Knockout," in *Methods in Gene Biotechnology*, pages 339-365 (CRC Press 1997)).

10

15

## CLAIMS

What is claimed is:

1. An isolated polypeptide, comprising amino acid residues 16 to 185 of SEQ ID NO:2.
2. The isolated polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.
3. An isolated nucleic acid molecule, wherein the nucleic acid molecule encodes amino acid residues 16 to 185 of SEQ ID NO:2.
4. The isolated nucleic acid molecule of claim 3, comprising the nucleotide sequence of nucleotides 46 to 555 of SEQ ID NO:1.
5. The isolated nucleic acid molecule of claim 4, wherein the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO:1.
6. A vector, comprising the isolated nucleic acid molecule of claim 3.
7. An expression vector, comprising the isolated nucleic acid molecule of claim 3, a transcription promoter, and a transcription terminator, wherein the promoter is operably linked with the nucleic acid molecule, and wherein the nucleic acid molecule is operably linked with the transcription terminator.
8. A recombinant host cell comprising the expression vector of claim 7, wherein the host cell is selected from the group consisting of bacterium, yeast cell, fungal cell, insect cell, avian cell, mammalian cell, and plant cell.
9. A method of using the expression vector of claim 7 to produce a polypeptide comprising amino acid residues 16 to 185 of SEQ ID NO:2, comprising culturing recombinant host cells that comprise the expression vector and that produce the polypeptide.
10. The method of claim 9, further comprising isolating the polypeptide from the cultured recombinant host cells.

11. An antibody or antibody fragment that specifically binds with the polypeptide of claim 1.
12. A composition, comprising a carrier and the polypeptide of claim 1.
13. A fusion protein, comprising the polypeptide of claim 1.
14. An isolated polypeptide, comprising an amino acid sequence selected from the group consisting of: amino acid residues 1 to 32 of SEQ ID NO:2, amino acid residues 16 to 32 of SEQ ID NO:2, amino acid residues 33 to 77 of SEQ ID NO:2, amino acid residues 79 to 102 of SEQ ID NO:2, amino acid residues 104 to 139 of SEQ ID NO:2, amino acid residues 141 to 173 of SEQ ID NO:2, and amino acid residues 175 to 185 of SEQ ID NO:2.

## SEQUENCE LISTING

<110> ZymoGenetics, Inc.

<120> Human Phermone Polypeptide

<130> 00-86PC

<160> 5

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 555

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(555)

<400> 1

atg atg ctg ctg ttg ctg tgt ctg ggg ttg acc ctc gtc tgt gcc cag	48
Met Met Leu Leu Leu Leu Cys Leu Gly Leu Thr Leu Val Cys Ala Gln	
1 5 10 15	

gag gaa gaa aac aat gat gct gtg aca agc aac ttc gat ctg tca aag	96
Glu Glu Glu Asn Asn Asp Ala Val Thr Ser Asn Phe Asp Leu Ser Lys	
20 25 30	

att tca gga gag tgg tat tcg gtt ctc ttg gcc tct gac tgc agg gaa	144
Ile Ser Gly Glu Trp Tyr Ser Val Leu Leu Ala Ser Asp Cys Arg Glu	
35 40 45	

aag ata gaa gaa gat gga agc atg agg gtt ttt gtc aaa cac att gat	192
Lys Ile Glu Glu Asp Gly Ser Met Arg Val Phe Val Lys His Ile Asp	
50 55 60	

tac ctg ggg aat tct tct ctg act ttt aaa ttg cat gaa att gaa aat	240
Tyr Leu Gly Asn Ser Ser Leu Thr Phe Lys Leu His Glu Ile Glu Asn	
65 70 75 80	

gga aac tgt act gaa att aat ttg gct tgt aaa cca aca gaa aag aat 288  
 Gly Asn Cys Thr Glu Ile Asn Leu Ala Cys Lys Pro Thr Glu Lys Asn  
                     85                    90                    95

gcc ata tgt agt act gac tat aac gga ctt aat gtc att gac ata ctt 336  
 Ala Ile Cys Ser Thr Asp Tyr Asn Gly Leu Asn Val Ile Asp Ile Leu  
                     100                    105                    110

gaa acg gac tat gat aat tat ata tat ttt tat aac aag aat atc aag 384  
 Glu Thr Asp Tyr Asp Asn Tyr Ile Tyr Phe Tyr Asn Lys Asn Ile Lys  
                     115                    120                    125

aat ggg gaa aca ttc cta atg ctg gag ctc tat gtt cga aca ccg gat 432  
 Asn Gly Glu Thr Phe Leu Met Leu Glu Leu Tyr Val Arg Thr Pro Asp  
                     130                    135                    140

gtg agc tca caa ctc aag gag agg ttt gtg aaa tat tgt gaa gaa cat 480  
 Val Ser Ser Gln Leu Lys Glu Arg Phe Val Lys Tyr Cys Glu Glu His  
                     145                    150                    155                    160

ggg att gat aag gaa aac ata ttt gac ttg acc aaa gtt gat cgc tgt 528  
 Gly Ile Asp Lys Glu Asn Ile Phe Asp Leu Thr Lys Val Asp Arg Cys  
                     165                    170                    175

ctc cag gcc cga gat gag gga gca gcc 555  
 Leu Gln Ala Arg Asp Glu Gly Ala Ala  
                     180                    185

<210> 2

<211> 185

<212> PRT

<213> Homo sapiens

<400> 2

Met Met Leu Leu Leu Leu Cys Leu Gly Leu Thr Leu Val Cys Ala Gln  
   1                    5                    10                    15  
 Glu Glu Glu Asn Asn Asp Ala Val Thr Ser Asn Phe Asp Leu Ser Lys  
                     20                    25                    30  
 Ile Ser Gly Glu Trp Tyr Ser Val Leu Leu Ala Ser Asp Cys Arg Glu  
                     35                    40                    45  
 Lys Ile Glu Glu Asp Gly Ser Met Arg Val Phe Val Lys His Ile Asp  
   50                    55                    60

Tyr Leu Gly Asn Ser Ser Leu Thr Phe Lys Leu His Glu Ile Glu Asn  
 65 70 75 80  
 Gly Asn Cys Thr Glu Ile Asn Leu Ala Cys Lys Pro Thr Glu Lys Asn  
 85 90 95  
 Ala Ile Cys Ser Thr Asp Tyr Asn Gly Leu Asn Val Ile Asp Ile Leu  
 100 105 110  
 Glu Thr Asp Tyr Asp Asn Tyr Ile Tyr Phe Tyr Asn Lys Asn Ile Lys  
 115 120 125  
 Asn Gly Glu Thr Phe Leu Met Leu Glu Leu Tyr Val Arg Thr Pro Asp  
 130 135 140  
 Val Ser Ser Gln Leu Lys Glu Arg Phe Val Lys Tyr Cys Glu Glu His  
 145 150 155 160  
 Gly Ile Asp Lys Glu Asn Ile Phe Asp Leu Thr Lys Val Asp Arg Cys  
 165 170 175  
 Leu Gln Ala Arg Asp Glu Gly Ala Ala  
 180 185

<210> 3

<211> 555

<212> DNA

<213> Artificial Sequence

<220>

<223> This degenerate nucleotide sequence encodes the amino acid sequence of SEQ ID NO:2.

<221> misc\_feature

<222> 9, 12, 15, 18, 24, 27, 30, 33, 36, 39, 45, 69, 72, 75, 78, 90, 93, 102, 105, 117, 120, 123, 126, 129, 132, 141, 162,

165, 171, 174, 180, 198, 201, 207, 210, 213, 216, 225, 243,

252, 264, 267, 276, 279, 291, 300, 303, 315, 318, 324

<223> n = A,T,C or G

<221> misc\_feature

<222> 336, 342, 390, 396, 402, 408, 414, 420, 423, 426, 429, 435, 438, 441, 447, 456, 462, 483, 510, 513, 519, 525, 531, 537,

540, 549, 552, 555

<223> n = A,T,C or G

<400> 3

atgatgytny tnytnytntg yytnggnytn acnytngtnt gygcncarga rgargaraay 60  
 aaygaygcng tnacnwsnaa ytygayytn wsnaarathw sngnggartg gtaywsngtn 120  
 ytnytngcnw sngaytgymg ngaraarath gargargayg gnwsnatgmg ngtnnttygtn 180



aarcayathg aytayytngg naaywsnwsn ytnacnttya arytncaayga rathgaraay 240  
 ggnaaytgya cngarathaa yytngcntgy aarccnacng araaraaygc nathtgywsn 300  
 acngaytaya ayggnytnaa ygtathgay athytngara cngaytayga yaaytayath 360  
 tayttytaya ayaaraayat haaraayggn garacnttyy tnatgytnga rytntaygtn 420  
 mgnacnccng aygtwnsnws ncarytnaar garmgnttyg tnaartaytg ygargarcay 480  
 ggnathgaya argaraayat httygayytn acnaargtng aymgntgyyt ncargcnmgn 540  
 gaygarggng cngcn 555

<210> 4

<211> 16

<212> PRT

<213> Artificial Sequence

<220>

<223> Peptide linker.

<400> 4

Gly Gly Ser Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser  
 1 5 10 15

<210> 5

<211> 6942

<212> DNA

<213> Homo sapiens

<400> 5

cagtgactga acatcggttg agcatgaacc aggaaagctc acatgagggt ctgggatgac 60  
 tttttaagca gtctacagac agagtgtcac cattcctacg ttcatgccc tgtgctcagc 120  
 tcctcccatg atgctaaact ctttctagtc ttgggggtgc agaagaagac cattagcaca 180  
 aggcagaaat gagctatata acaaccctgg ctggtggcat gagatcttat ctcagtatat 240  
 tcctccctcc tggaatgcta ttcccacatg cactccatat ctccataccc tccctgactt 300  
 tcaacaccta ctacaagtcc taccttatcc gggaagcatt ctctgaccca ccagctcaat 360  
 gtgtgctaga catgatgctg ggggttttac gcacatcacc ccattgaatc cctagcagtg 420  
 tattaacat gtcattagcc ctgttaagag gaactaacag gctgggtgtg gtggctcatg 480  
 cctgtaatcc cagcactttg ggaggccaag gcagggtgat cacgagggtca agagatggag 540  
 gccatcctgg ccaacatggt gaaaccccggt ctctcctaaa aatacaaaaat tagccaggag 600  
 tgggtggcaca tgcctacagt cccagctact cgggaggctg agggaggagaa tcacttgaac 660  
 ccaggaggca gaggttgacg tgagccgaga tcacaccatt gccccccagc ctggtgacag 720  
 agcgagattc catctcaaaa aaaaaaaaaa agaggaaact aacagagaga aggctcagag 780  
 agaagtcaca gtctgtcttt atggcttcaa aacagcagag ataggaatga agcccaggtc 840  
 ttcagtactc caaaacctgt gttcttaggc tcttttgaca tgatgcccc actccagccc 900  
 cttttcaatt gtaccatggg aaaagttgac agggcttggg cacctccatt attaataataa 960  
 acctttatta aagtttataa tagtgacttt attagggttac tcaattgggt catttcatgt 1020

gcttgtaggt ttaaattttt actcagtatg tgaaacattt ttccaacttc actaaagggt 1080  
 ctttgaagga aggaatgata tcttgtactt ctccttttct cttatgggcc catgtgtatg 1140  
 tctcttatgg taaagacata cacatagtga ctaaattttg ttcacttatt gtggggagag 1200  
 ggaaggctgc aggggttggg ttcagagaat aaggaaaagg aatttgaatg tactgaaaac 1260  
 cagacaccat actttaaaga ctttttctgt tcaaccctat tttaatgctt tgtgataggt 1320  
 gctttcatct tcggtagaac agtctgtttc tggggcttgt aaaaattgag taactcacct 1380  
 aatttcacac attgtatgag aagtagaaca gagattctgg cacatcctgt ttctcagtta 1440  
 tgcccatcc tgccacacat aatcggtctt ctgatttttag gccagatcca cagttaaatac 1500  
 tagttccaaa tgacagcccc ttagagcatc tgctattttt cttgataatg tgttctggga 1560  
 aatttgaggt cactaactag atctaagga gacggaattg ccaagtttca aaagggcagg 1620  
 aacaatcctt ggttgacacac cagtgaagga aagaccattt catggagggg gaagggaaag 1680  
 gtagtaccac catgggcaaa acactggaca cagactggat ataaagacag atgagctggg 1740  
 gagtggagcc cactgctaga gaaagacca tcccagcaa ctgtggagga ggcagtgtg 1800  
 tcccttacca agatgatgct gctgttgctg tgtctgggtg tgaccctcgt ctgtgccag 1860  
 gaggaagaaa acaatgatgc tgtgacaagc aacttcgatc tgtcaaagggt agagttatgg 1920  
 aacgcattga cttctggctc tggggagtgg tttcagggat tctgaagcca tgtggatcct 1980  
 attgtagggc aggagttcct taacatatca gacaaaagca ttggtttctg atcttgggtt 2040  
 aagagcaatt gtaccatatg aaaagatgac aatgcttggg gatctccatc ctccctaaag 2100  
 gggcagtggg aagagcccta ggtgtggtgc caactgatct gagttctatc ttcaactcac 2160  
 cattatgagc cctggcccac aggcagagt tcaaggaaag aagtgtccat gtggtatccc 2220  
 cactcctctt ttcttaccac atgaccttat ttattgcaga tttcaggaga gtggtattcg 2280  
 gttctcttgg cctctgactg cagggaaaag atagaagaag atggaagcat gagggttttt 2340  
 gtcaaacaca ttgattacct ggggaattct tctctgactt ttaaattgca tgaaatataa 2400  
 gtatggcctt ttttagttgg ggaaggaaaa acagaagcct gggatatggc ttgaacacac 2460  
 acacattcac tgatccagga cttgatggat ctggggcttt tcaagattca gagacaagag 2520  
 tcagtaacaa gaatttttaa ttttaattgga gtgggagaag caggaggtag agggatatca 2580  
 gaatcattaa gtagtatgac aatatagcat aggttaagca aactgaattt ggaactaaca 2640  
 gcctgaattt gaatcttcca gttaccagtt gtgttaccac gaaaactaat gtctttccac 2700  
 ctcccttttc ccatgggtaa aaggggatga caataattcc tcctcatgtt gttttgtttt 2760  
 gaggtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtttgtgtg gaccaaggca 2820  
 ggtcacattt gtaaagcact tacttgagag gctggcatgt ggtgaacact cagctaataga 2880  
 tctagacaag tagttgtagt aatagcaaca gtgagaagta gcaacagtga tagaagttgt 2940  
 catttaagtc attcttatca aatcatgtga aataatacac ataaaagggtc ttgtataatt 3000  
 gtacaaaaaa atgttcaaaa gttgaatcag aaactcaagt tactggtaga ggagaaggta 3060  
 agcaagccag cctcatcagt agcccttggg catcaattta ttgttttaca gtgaaaatgg 3120  
 aaactgtact gaaattaatt tggcttgtaa accaacagaa aagaatgccat tatgtagtac 3180  
 tgactgtaag tattttttta agtcttttcta aattatactt ccaactgcaa ggaattttact 3240  
 aaacccagat atcttcatga tcttgatcaa tttctgatca atgactatct tgataataaa 3300  
 atgggggtta tataatagcc aataacctga agctgggaga accttgcatg taaaagccag 3360  
 tgcctgccac ataatacaat cagcttcagg aaagtgaact ggtcatatgc acctcctgca 3420  
 aaaaaagtta attggcaggg aaaaaatata aattggggct gggcatggta gctcacacct 3480  
 gtaatcccg cactttggga ggccaaggta gggggatcac ctgaggtcag gagttcgaga 3540  
 ccagcctgac caacatggtg aaaccctgtc tctactaaaa atacaaaaat tagctgggca 3600

tcatgatgca cacctgtgat ctcagctact cgggaggctg aggcaggaga atcgcttgaa 3660  
ccgtgaagga ggaggttgca gtgagccaag attgcgccac tgcactccag cctgggcaac 3720  
aaagcgagac cttatcttta aaaaaaaaaa aaatggtaaa ttgtcaagac agcttcttat 3780  
tccctccctg gctttctgat gcctgatggt catcgccaca gcacattgct caggatatag 3840  
catcaagtaa gtaatgccag tgggggctta tagagaggac agagacacat cactgacaga 3900  
gatgctgctg gaggtgggta ggtctcgggg gggatgttaa agtctgagt gctgctgggg 3960  
ctaaattttt tgactgctag tggtaacct acccttttct cagaatttgc ctcataagc 4020  
ccagggcgat gcaaattgaa gacattcaac atcctcacta tctctaacag gatacgact 4080  
tgaacttggt ttgtgggaga gaccaggtaa aggggttcaa ctgctttcta tggaccaaag 4140  
gagttctgag ctgtgttggt tcctccacag ataacggact taatgtcatt gacatacttg 4200  
aaacggacta tgataattat atatattttt ataacaagaa tatcaagaat ggggaaacat 4260  
tcctaagtct ggagctctat ggtagagtat tttcatgctg acactcatct tggggtgggg 4320  
aaaggaggga aggcaagaaa cttagatata aagtggctct tcccacgtta tcctggggcca 4380  
tctctccttc tctgcaaaca gaggcttacc tttggtaatc cctcaagctc agtgataaga 4440  
aattatgac ttccattgtg ctactgatgt caaatttagg acacattata tcccaccccc 4500  
aagacttctg cctggcatcc aatggctgcc attcattttc ccatctgtct tgagccctgt 4560  
cccctcattt gcctctatga cgaataccca agacatggca gttactgtca ctcttcccac 4620  
agttcgaaca ccggatgtga gctcacaact caaggagagg tttgtgaaat attgtgaaga 4680  
acatgggatt gataaggaaa acatatttga cttgaccaa gttggtaagt cggggtttct 4740  
ggtattctct tcctaaattc ccatgttaca gaaggagca atccagggga agtaataccc 4800  
tccaaatcag tcccctttgt gctaacttg agaatcatct ggtttctctc tgaaaaataa 4860  
actgggggtg gtgtctgaat ttgttctctt ctccctcgac agatcgctgt ctccaggccc 4920  
gagatgagg agcagcctag gactccgggt tggatgctc tgacaccggt ggagagagg 4980  
tggcccagg accagtgcct tccaaaagca ttaggggtt gcacccaaag ataccataa 5040  
aataatttg taggaaagct tgtgggaaaa tcttgaatc tggagttgga agacctggat 5100  
agaggacca atttctccat aagctttgag caaataacct aagacttaa aaggagacaa 5160  
tgggccagg acagtggctc acacctgtaa tcccagcact tagggaggca gtggcaggag 5220  
gatcgctga gcccaggagc ccaagaccag cctgggcaac ataatgagat accattctcc 5280  
atgaaagga gaaaaaaaaa agaagaggaa acaatgatgc ctagtttaca gagttgctgt 5340  
gtagatataa tgcttgcgaa aaggggccag taattgaaga tgtgctaaaa atcactagga 5400  
tttcacgttc gaaatggaaa aagtagtagg ttgtatgaaa ctcaaataatg tttctgtgtt 5460  
tgtggacca ctcccaggag gccatctcac cctccaactt ataatccaca tttcttcttc 5520  
tgtactctgt atgtccttga cgtctaacct gggaaatggaa ctgtcttccc attttccaca 5580  
aacctggaaa atagatcttg agttttctca actaccttag atttatcgtg gggctttcag 5640  
ataatgtatg tgaaagtctt agcagcttgc ataaagcaaa tatttaataa gagatatcta 5700  
cagattttgc cattattcct caatttatca aaaattcttc aaaacttctg atgttgaaga 5760  
tggcagaaac aaacctggga ccctttatta cttcctttct ctttttcatg tattcagctg 5820  
agagaagggt aagatggcat tagcaatagc aaaaaagcaa aagtcattta ctcaacatcc 5880  
tccctcccc ttttctggta tacaacataa gttagaatac agggtcagaa ctatttcttc 5940  
atcatttttc tctgtaatag aatttctctt catcatcat gcaatgcaag gaatgtttaa 6000  
taagttcctg tgttctctct cttctctgct taattagtgc tgagtgaatg agtgatgagt 6060  
gaatgagtga atgctgagt ctgagtgaat gtttcctcac ctggactcca gtatcttccc 6120  
ttcctgatcc ccatgtcctc ttatgacaag ttctgtgacc tgatttccat catacacac 6180

atgaagacgt catccctgca tctttaggat cttccctaac tgcctaagaa gactcagaaa 6240  
ttcaccaaga atcaaagggtt ttctttgaat ttctgactgt ctttgtcaca gccaggggaa 6300  
ctctacatga ataacacctg actctacatg atcaataaat cattaacctt gcagtcatgt 6360  
aatttgtctt tgtatctgtg aggatgtgag gctgacagtg agtgaatgaa ggggcttcct 6420  
tgcccctgaa cccataatcc tcctctagta tctcagaatc ttctctgagc ttcaggattc 6480  
tccttcctct gcagtaattc cctggttggtt atcccaagtc cccatgagtg ctgcttgccc 6540  
cttgatttgc caatttctta gtgatctgat aagaacatta taaataatat aagaatataa 6600  
gacttgcccc atgatttctg catcctagaa ttcgcttttc aattggggac actgacagat 6660  
agatgctgct gaggctgata ccagcttgac agagatgcct ctagatacta gcttctctat 6720  
accatccatc ctataagatg gggttcagtgt cttttcacac tcttgaaaaa ccccttcaa 6780  
tcaccacagt tatctactgt gaacacagta agtgagattt cattaaatgc aaacaatacc 6840  
tgtcttcagc aacctgtact tggttttgtc aggaggcact gatggggaat gttattaact 6900  
tcttcttttg aaagctaaca ggcttgggca aagggatgtg aa 6942